MISURE DI GRANDEZZE FISICHE

IMPORTANZA DELLE INCERTEZZE NELLE MISURE FISICHE

Nell'ambito delle scienze sperimentali la parola *errore* non ha la tradizionale connotazione di *equivoco* o di *sbaglio*, bensì assume il significato di **incertezza** intesa come quell'entità di cui sono affette inevitabilmente tutte le misure.
Infatti, se si potessero trattare le incertezze sperimentali alla stregua di veri e propri errori, basterebbe agire sulle fonti di questi o sforzarsi di operare nel modo più accurato possibile per eliminarli, mentre è inevitabile che esse siano presenti nel processo di misura.

Da questa premessa dobbiamo allora concludere che tutto quello che eseguiamo all'atto della misura e affetto da errore?
Se così è, che significatività possiamo attribuire ai nostri risultati?
Ovviamente i risultati che otteniamo non possono essere assunti come certezze assolute, né si possono sfruttare per ottenere previsioni corrette al 100% : nonostante tutto entro i limiti di validità dei nostri dati (limiti dati appunto dagli errori ad essi associati) possiamo ugualmente fare delle previsioni sull'evolversi dei fenomeni che stiamo studiando.
Si noti comunque che l'intervallo di validità determinato dalle incertezze sperimentali non fissa una linea di demarcazione netta tra quello che è il risultato corretto sul quale costruire le nostre congetture e quello che invece può essere un valore completamente inconsistente: diciamo che all'interno del cosiddetto *intervallo di confidenza* siamo "più sicuri" che altrove di avere inglobato quello che è il valore effettivo della grandezza che misuriamo.

In definitiva cercare di ottenere una stima quantomeno realistica della loro consistenza e, per quanto possibile, sforzarsi di ridurle ad un limite ragionevole è quanto di meglio si possa compiere in una "corretta analisi" delle incertezze sperimentali.
Il significato delle virgolette è da ricercarsi nel fatto che l'analisi degli incertezze non costituisce una vera e propria *teoria* degli errori (come a volte si trova scritto impropriamente su alcuni testi più o meno scientificamente accreditati), bensì dovrebbe limitarsi a dare alcune linee guida fondamentali per l'analisi e il trattamento dei dati sperimentali.

GLI ERRORI NELLE SCIENZE SPERIMENTALI
- Un limite ragionevole -

Una vera e propria definizione di limite ragionevole per le incertezze sperimentali non esiste, né, suppongo, saremmo in grado di coniarne una seduta stante: infatti molto spesso il limite oltre il quale l'incertezza associata al valore cercato viene ritenuta accettabile non è ben definito, o non lo è affatto.
I criteri secondo i quali un intervallo di incertezza viene considerato significativo o meno possono

essere dettati da molteplici fattori tra cui la precisione richiesta nell'esperimento o l'esperienza e la familiarità con i fenomeni in esame maturata dallo sperimentatore.

In altre parole, "...non vi è alcuno scopo nel pesare una carrata di fieno sulla bilancia di precisione del chimico....": questa frase di Oskar Anderson riassume nella sua semplicità e immediatezza quanto detto.
E' inutile ad esempio tentare di misurare con la precisione del milionesimo di millimetro, magari usando attrezzature costose con un notevole dispendio di tempo, la lunghezza di un mobile per vedere se entra nell'angolo della nostra casa predisposto ad accoglierlo: basterà infatti una misura effettuata con un metro la cui precisione è dell'ordine del mezzo centimetro o al limite del millimetro per soddisfare la nostra richiesta di adattabilità o meno del mobile al sito predisposto.

Un bravo sperimentatore dovrà essere quindi in grado di valutare quale sia l'ordine di grandezza della precisione richiesta e operare di conseguenza, sia a livello di scelta degli strumenti adatti sia come tecniche di lavoro, "ottimizzando" il proprio lavoro.
L'abilità nello scegliere la strada giusta da percorrere per affrontare un esperimento e le relative misure la si matura soprattutto con l'esperienza diretta nei laboratori: da qui l'impossibilita di prescindere da esso nella formazione di uno studente nonché di un futuro ricercatore.

IMPOSSIBILITÀ DI OTTENERE MISURE FISICHE PRIVE DI INCERTEZZE

Relativamente al "valore vero" bisogna specificare che esiste un problema di ***definizione*** del concetto stesso.

Se ad esempio supponiamo di voler misurare la lunghezza di un tavolo con una precisione sempre maggiore arriveremo ben presto a renderci conto che non esiste la grandezza "lunghezza del tavolo": infatti all'aumentare della precisione noteremmo che la misura della lunghezza varia da punto a punto a causa di piccole asperità del bordo del tavolo, si differenzia a diversi intervalli di tempo per la dilatazione o la contrazione dovuta agli sbalzi termici e via dicendo....
L'esempio mostrato illustra in definitiva il seguente fatto: *nessuna quantità fisica può essere misurata con completa certezza.*
Pur sforzandoci di operare con la massima cura non riusciremo mai ad eliminare totalmente le incertezze: potremo solo ridurle fino a che non siano estremamente piccole, ma mai nulle.

Da questo si capisce il perchè dell'affermazione "non esiste la grandezza lunghezza del tavolo" intesa come valore vero: essendo infatti ogni grandezza definita dalle operazioni sperimentali che si svolgono e dalle regole che si applicano per ottenere la misura, il fatto di non poter eliminare del tutto le incertezze e, al di sopra di una certa soglia di precisione, trovare, al ripetere della misura, valori diversi, nega l'esistenza di tale grandezza.

Per capire meglio, il valore vero sarebbe il risultato di operazione di misura ideale, priva di errore: tale misura nella realtà è irrealizzabile, pertanto anche il relativo risultato, il valore vero, perde di significato.
Infatti, come nella realtà affermiamo che qualcosa esiste solo nel momento in cui i nostri sensi sono in grado di determinarne la presenza, cosi nel caso delle misure, noi possiamo affermare che la grandezza "lunghezza della porta" esiste solo quando siamo in grado di definirla esattamente.

La presenza inevitabile di incertezze nella misura elimina automaticamente la possibilità di determinare esattamente tale grandezza.

IL DETERMINISMO

L'ideale deterministico, tipico della cultura dell'era newtoniana, fu generalizzato in modo formale da Laplace nell'introduzione della sua opera *"Essai philosophique sur les probabilites"*.

Secondo il determinismo ogni fenomeno della natura doveva essere considerato nell'ambito di una logica rigorosa connessa al presupposto che ogni evento fosse ricollegabile ad una causa che lo provocava.
Indotti da questo formalismo, molti scienziati erano convinti che una volta conosciuto lo stato iniziale di un sistema e le forze agenti su di esso fosse possibile determinare istante per istante l'evolversi del sistema applicando le leggi della meccanica di Newton.

La "fiducia" nelle potenzialità della scienza e della ricerca derivavano in larga misura dal ritenere concettualmente possibile conoscere la posizione e la velocità istantanea di tutte le componenti dell'universo: si pensava cioè che fosse possibile, almeno in via di principio, affinare le tecniche e i metodi di misura in modo da far scomparire definitivamente le indeterminazioni sui valori misurati.

Al giorno d'oggi questa visione è stata completamente abbandonata in quanto minata dal principio di indeterminazione di Heisenberg e rimpiazzata dalla visione probabilistica introdotta dalla meccanica quantistica.

IL DETERMINISMO DI LAPLACE

Questa frase di Pierre Simon de Laplace (filosofo e matematico francese 1749-1827), tratta dalla sua opera *"Essai philosophique sur les probabilites"*, riassume in modo significativo la visione del determinismo secondo Laplace:

"Noi dobbiamo riguardare il presente stato dell'universo come l'effetto del suo stato precedente e come la causa di quello che seguirà. Ammesso per un istante che una mente possa tener conto di tutte le forze che animano la natura, assieme alla rispettiva situazione degli esseri che la compongono, se tale mente fosse sufficientemente vasta da poter sottoporre questi dati ad analisi, essa abbraccerebbe nella stessa formula i moti dei corpi più grandi dell'universo assieme a quelli degli atomi più leggeri. Per essa niente sarebbe incerto ed il futuro, così come il passato, sarebbe presente ai suoi occhi."

MISURE: METODI E STRUMENTI

LE CARATTERISTICHE DEGLI STRUMENTI DI MISURA

- Ripetibilità
- Prontezza
- Sensibilità
- Risoluzione
- Fondo scala
- Precisione

Oltre a queste esistono altri fattori tra cui il costo, l'ingombro e il peso che contraddistinguono lo strumento: si noti peraltro che le varie caratteristiche non sono indipendenti l'una dall'altra, ma costituiscono il risultato di un compromesso che si raggiunge all'atto della progettazione.
Come esempio si consideri il fatto che il costo di uno strumento può salire notevolmente all'aumentare dell'intervallo di funzionamento e non è detto che le due grandezze, costo e intervallo di funzionamento, siano correlate linearmente: anzi molto spesso ad un piccolo ampliamento delle caratteristiche (precisione, intervallo di funzionamento e così via) corrisponde, in proporzione, un aumento dei costi ben superiore ai miglioramenti apportati.

LA RIPETIBILITA`

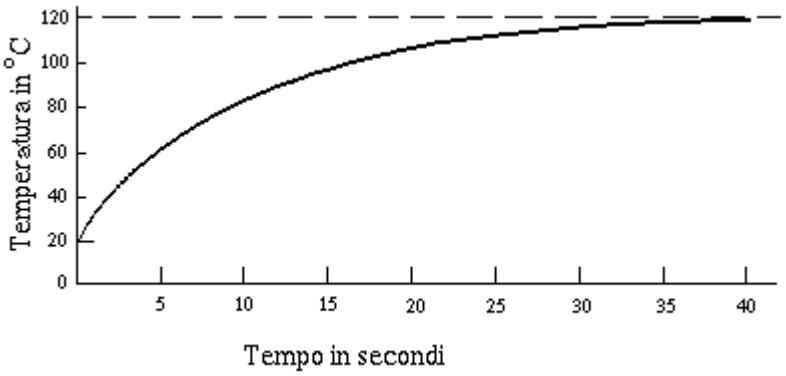
Con il termine ripetibilità si intende la capacità dello strumento di fornire misure uguali della stessa grandezza entro la sua risoluzione, anche in condizioni di lavoro difficili o variabili (vibrazioni, sbalzi di temperatura, ...).
In pratica lo strumento deve risultare ben isolato rispetto agli effetti dell'ambiente esterno, escluso ovviamente l'effetto dovuto alla grandezza in esame.

La ripetibilità implica anche una buona **affidabilità** intesa come robustezza di funzionamento nel tempo: questa peculiarità viene espressa come *vita media* o come *tempo medio* statisticamente prevedibile fra due guasti successivi in condizioni normali di utilizzo.

LA PRONTEZZA

La prontezza è una caratteristica dello strumento legata al tempo necessario affinché questo risponda ad una variazione della grandezza in esame. Per alcuni, quanto minore è questo tempo, detto *tempo caratteristico*, tanto maggiore è la prontezza, mentre per altri la prontezza è rappresentata dal tempo impiegato dallo strumento per dare la risposta, cioè il risultato.
In generale la prontezza rappresenta *la rapidità con cui è lo strumento è in grado di fornire il risultato di una misura.*

Per chiarire quanto detto finora vediamo un esempio: consideriamo un termometro a mercurio, quello che si può trovare in un comune laboratorio, che sia inizialmente alla temperatura ambiente di 20^{o}C.
Se ora lo immergiamo in un bagno di liquido alla temperatura di 120^{o}C osserviamo che il mercurio comincia a salire lungo la scala prima velocemente poi più lentamente fino ad arrivare al valore di temperatura corrispondente: approssimativamente il tempo impiegato affinché il mercurio raggiunga la posizione relativa alla temperatura misurata è dell'ordine di qualche decina di secondi (diciamo 40).
Questo intervallo di tempo ci da un'indicazione sulla prontezza dello strumento: in particolare se andiamo ad osservare l'andamento della temperatura misurata graficata rispetto al tempo, il fenomeno descritto appare ancora più chiaro.



C'è anche chi definisce la prontezza come il tempo impiegato dall'indice dello strumento (nel nostro caso il livello della colonnina di mercurio) ad effettuare il 63.2 % dell'escursione che esso deve compiere, partendo dalla posizione iniziale di riposo fino a raggiungere il valore effettivo della grandezza: tale tempo è definito come *coefficiente di ritardo*.
Attraverso questa definizione si potrebbe avere un coefficiente di ritardo variabile con il valore della grandezza applicata: per ovviare a questo inconveniente occorre fissare un valore di riferimento della grandezza, le modalità d'uso e tutte le altre caratteristiche strumentali in modo tale che la prontezza così definita rispecchi un'effettiva caratteristica dell'apparecchio.

LA SENSIBILITA`

La **sensibilità** di uno strumento è costituita dalla più piccola grandezza in grado di generare uno spostamento apprezzabile rispetto all'inizio della scala dello strumento.
Così definita, la sensibilità determina il limite inferiore del campo di misura dello strumento, mentre il limite superiore è dato dal fondo scala: i due determinano insieme l'*intervallo di funzionamento*.

Esiste anche una definizione più raffinata di quella presentata, anche se concettualmente sono del tutto equivalenti.

LA RISOLUZIONE

La risoluzione di uno strumento rappresenta la minima variazione apprezzabile della grandezza in esame attraverso *tutto* il campo di misura: essa rappresenta il valore dell'ultima cifra significativa ottenibile.
Per cui se la scala dello strumento parte da zero ed è lineare la risoluzione è costante lungo tutto il campo di misura e risulta numericamente uguale alla sensibilità.

Si osservi che non sempre sensibilità e risoluzione coincidono: la loro differenza risiede nella definizione delle due grandezze. Infatti la sensibilità è relativa all'inizio del campo di misura, mentre la risoluzione è considerata sull'intero campo di misura dello strumento.

IL FONDO SCALA

Il fondo scala rappresenta il limite superiore del campo di misura e prende anche il nome di *portata* dello strumento: insieme alla sensibilità ne delimita l'*intervallo di funzionamento*.

LA PRECISIONE

Come abbiamo già detto, ad ogni misura è associata inevitabilmente una incertezza. Evidentemente più piccola è l'incertezza associata alla misura, migliore sarà la misura.
Ma cosa significa "più piccola"?

Vediamo di chiarire questo punto. Quando noi forniamo un risultato, lo dobbiamo sempre corredare, oltre che del valore della misura, anche dell'errore associato: tale errore è detto *errore assoluto* e rappresenta l'intervallo di indeterminazione entro il quale si suppone che il risultato sia compreso.
Se ora consideriamo il rapporto tra l'errore assoluto e il risultato stesso otteniamo una grandezza adimensionale (un numero, privo cioè di unità di misura), molto utile nell'analisi degli errori, che prende il nome di **precisione** o errore relativo.

A questo punto appare evidente che la misura con l'errore relativo minore è quella più precisa: si noti bene che si è parlato di errore relativo e non assoluto. Infatti si consideri il seguente esempio.

Siano date due misure nel modo seguente

$A=(10 \pm 1)$ Kg

$B=(100 \pm 1)$ Kg

Entrambe hanno lo *stesso errore assoluto* ($\Delta A=\Delta B=1$ Kg), mentre hanno *differenti errori relativi*. Ora, mentre nella prima misura abbiamo un errore di una unità su dieci, nella seconda abbiamo un errore di una sola unità su cento: si è allora soliti dire che la prima è una misura *precisa al 10%*, mentre la seconda *precisa al 1%*.

Precisioni di questo ordine di grandezza sono molto simili a quelle che si possono ottenere in un laboratorio di fisica o di chimica: si tenga però conto che i laboratori di ricerca le precisioni raggiunte sono di parecchi ordini di grandezza superiori. Per questo si è soliti usare la notazione scientifica, onde evitare la scomodità di espressioni con troppi zeri.

COMPONENTI FONDAMENTALI DEGLI STRUMENTI DI MISURA

- Elemento rivelatore
- Trasduttore
- Dispositivo per la visualizzazione

ELEMENTO RIVELATORE

L'elemento rivelatore dello strumento è costituito da un apparato sensibile alla grandezza da misurare: in un termometro, ad esempio, l'elemento sensibile è rappresentato dal mercurio.
in generale l'elemento rivelatore interagisce con la grandezza in esame e, come avviene per il mercurio, può venire modificato nella forma o in un'altra caratteristica: altre volte invece è l'elemento stesso che, interagendo con la grandezza in esame, la modifica, creando non pochi problemi allo sperimentatore che deve accorgersi della modificazione introdotta e di conseguenza cercare di eliminarla o stimarla per ottenere una misura corretta.

TRASDUTTORE

Il trasduttore è quella parte dello strumento atta a trasformare l'informazione ottenuta dal rivelatore in una grandezza di più facile utilizzazione da parte dello sperimentatore. Il trasduttore in genere agisce sulla grandezza di partenza trasformandola in una di un'altra specie: nel caso di oscilloscopio il trasduttore è costituito dalle placchette di deflessione del fascio di elettroni, mentre per gli strumenti digitali, tale componente è rappresentato dal convertitore analogico-digitale.

DISPOSITIVO DI VISUALIZZAZIONE

Questo componente ha la scopo di fornire visivamente o graficamente il risultato della misura sintetizzando così le operazioni svolte dal rivelatore e dal transistore.
Ad esempio sono dispositivi di visualizzazione lo schermo per un oscilloscopio, l'insieme dell'ago e della scala graduata per un generico strumento ad ago mentre per gli strumenti di tipo digitale è in genere costituito dal display numerico.

GLI STRUMENTI DI MISURA ANALOGICI

La caratteristica primaria di uno strumento analogico, che lo differenzia da uno digitale, è il fatto che in esso è assente la fase di digitalizzazione del segnale in ingresso.
Infatti, secondo la definizione in ambito tecnico della parola, con "analogico" si intende un *dispositivo operante su grandezze fisiche elettriche, meccaniche o di altra natura che rappresentano, per* ***analogia****, le grandezze caratteristiche del sistema studiato,* ossia *mettono in relazione due fenomeni fisici di natura diversa in modo che le grandezze relative all'uno e all'altro siano legate da equazioni identiche:* in questo modo si possono studiare fenomeni complessi su modelli costituiti da fenomeni più semplici.

In pratica negli strumenti analogici non avviene la conversione del segnale in ingresso, ma questo viene studiato così com'è, sfruttando fenomeni e meccanismi interni allo strumento, analoghi alla natura del segnale.

Se volessimo riassumere lo schema di uno strumento analogico avremmo:

dove con ***G*** si intende la grandezza misurata, con ***M S*** i generici meccanismi interni dello strumento e con ***Vis*** il dispositivo di visualizzazione dello strumento: in genere, un indice mobile su una scala graduata.

GLI STRUMENTI DI MISURA DIGITALI

Per alcuni, la caratteristica principale che differenzia gli strumenti di tipo digitale da quelli analogici, è la presenza di un display a cristalli liquidi sul quale viene riprodotto il risultato come una sequenza di cifre.
In effetti, anche nel linguaggio di tutti i giorni, l'appellativo digitale spetta a tutti quei dispositivi che hanno un display a cristalli liquidi, mentre per analogico in genere si intende tutto ciò che è "a lancette". Questa distinzione non è del tutto errata, ciò non toglie che sia abbastanza riduttiva poiché non fa luce sulle differenze concettuali e di funzionamento dei due tipi i strumenti.

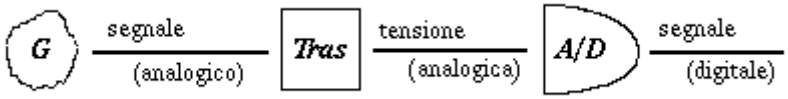
Fondamentalmente uno strumento digitale compie le seguenti operazioni:

- in ingresso legge un segnale, di tipo analogico, omogeneo alla grandezza osservata

- il trasduttore prende questo segnale e lo trasforma in una tensione, grandezza elettrica ed ancora analogica

- il segnale elettrico va al convertitore anologico-digitale che trasforma la tensione in ingresso in un segnale digitale.

I passi elencati si possono riassumere attraverso il seguente schema:



Una volta trasformato il segnale in ingresso da analogico a digitale lo strumento lo analizza a visualizza sul display il corrispondente risultato.
Vediamo in dettaglio come avvengono le due fasi principali appena citate e cioè come funziona un convertitore analogico-digitale e come viene elaborata l'informazione fornita da questo per ottenere il risultato visualizzato sul display.

LE MISURE
- Introduzione -

Il metodo di indagine sperimentale privilegiato per lo studio dei fenomeni naturali è quello sperimentale, applicato per la prima volta da Galileo Galilei (1564 - 1642).
Questo metodo si propone, dopo un'accurata osservazione, di riprodurre in condizioni semplificate e controllabili i processi che avvengono spontaneamente in natura: dopodiché si cerca di isolarne ogni singolo aspetto studiando l'influenza che questo ha sul fenomeno osservato.
Una volta individuati i fattori determinanti, si procede alle dovute valutazioni quantitative atte a sfociare nella formulazione matematica di una o più leggi.

Alla base di questo discorso risiedono due elementi fondamentali:

- La riproducibilità degli esperimenti
- La precisione delle misure

L'importanza della riproducibilità delle misure si intuisce immediatamente dal fatto che per poter studiare l'apporto dei singoli fattori bisogna poter ripetere le misurazioni dopo aver variato alcuni parametri del sistema e vedere quanto la variazione apportata incida sull'evolversi del fenomeno.
Ne segue, ed è il secondo punto, che l'affidabilità delle misure influenza notevolmente la qualità delle deduzioni che possiamo ricavare da un siffatto studio dei fenomeni: questo è dovuto al fatto che il campo di validità delle nostre previsioni, dedotte dall'osservazione sperimentale, è limitato dalla precisione con la quale effettuiamo le misure stesse.

MISURE
- La verifica degli strumenti -

Nel momento in cui si osserva un fenomeno che può essere difficilmente riprodotto le uniche fluttuazioni nelle misure alle quali bisogna prestare attenzione sono quelle legate all'apparato sperimentale.
Da questo discende immediatamente che l'affidabilità degli strumenti usati deve essere verificata attentamente prima di procedere al prelievo delle misure, operando in condizioni il più possibile simili a quelle reali dell'esperimento e, ovviamente, eseguendo misure su una grandezza omogenea a quella che è oggetto di studio.

Una volta verificati gli strumenti operando come sopra, si procede alla misurazione vera e propria agendo con la massima attenzione e nel modo più preciso possibile in quanto potremo raccogliere un unico dato. Essendo il dato uno solo, non avrà senso parlare di statistica nella futura analisi dei dati (...in questo caso **del** dato...), quindi non si potrà ad esempio usufruire degli espedienti statistici per ridurre l'errore associato: bisogna che la misura così come è stata prelevata sia la migliore possibile in quanto non sarà possibile raffinarla ulteriormente.

LE MISURE
- La discrepanza -

Nel momento in cui due misure sono in disaccordo diciamo che tra loro vi è una *discrepanza*. Numericamente definiamo la discrepanza come la *differenza tra due valori misurati della stessa grandezza*.
E` importante sottolineare che una discrepanza può essere o non essere significativa. Se due studenti misurano la capacità di un condensatore e ottengono i risultati

$$C_1 = (40 \pm 5) nF$$

e

$$C_1 = (42 \pm 8) nF$$

la differenza di 2 **nF** è minore dei loro errori: in questo modo le due misure sono ovviamente consistenti.
D'altra parte se i risultati fossero stati

$$C_1 = (350 \pm 20) nF$$

e

$$C_1 = (450 \pm 10) nF$$

allora le due misure sarebbero state chiaramente inconsistenti e la discrepanza di 100nF dovrebbe essere significativa. In questo caso si sarebbero resi necessari controlli accurati per scoprire gli errori commessi.

LE MISURE
- Confronto di due misure -

In molte esperienze si determinano due numeri che, in teoria, dovrebbero essere uguali. Per esempio, la legge di conservazione della quantità di moto stabilisce che la quantità di moto totale di un sistema isolato è costante.
Per provarlo, potremmo compiere una serie di esperimenti con due carrelli che si urtano quando si muovono lungo una rotaia priva di attrito. Potremmo misurare la quantità di moto totale dei due carrelli prima della collisione (***p***) e di nuovo dopo la collisione (***p'***), e verificare di seguito se ***p***=***p'*** entro gli errori sperimentali.
Per una singola coppia di misure, il nostro risultato potrebbe (ad es.) essere

$$p = 1.49 \pm 0.04 \ \ Kg \cdot m/s$$

e

$$p' = 1.56 \pm 0.06 \ \ Kg \cdot m/s$$

Qui l'intervallo in cui ***p*** probabilmente giace (1.45-1.53) si "sovrappone" all'intervallo in cui presumibilmente giace ***p'*** (1.50-1.62). Allora questa misura è consistente con la conservazione della quantità di moto.
Se, invece, i due intervalli probabili non fossero così vicini da sovrapporsi, la misura non sarebbe consistente con la conservazione della quantità di moto. A quel punto dovremmo verificare l'esistenza di errori, nelle misure o nei calcoli, la presenza di errori sistematici, o la possibilità che alcune forze esterne (come la gravità o l'attrito) abbiano causato la variazione della quantità di moto del sistema.

ANALISI DEGLI ERRORI DI MISURA

ERRORI
- Introduzione -

Ogni grandezza è definita dalle operazioni sperimentali che si compiono per ottenerne la misura: il risultato di una misura dipende perciò dal procedimento adottato e dalle caratteristiche dello strumento usato nonché da tutti quei fattori esterni che intervengono sullo svolgersi dell'esperimento e della misura.
Ora noi sappiamo che per quanto ci si possa sforzare di aumentare la sensibilità dei nostri strumenti o di affinare le tecniche di prelievo dei valori, tutte le misure che effettuiamo sono affette da errore: questo significa che esse non coincidono con il "valore vero" della grandezza misurata.
Ma cos'è che fa sì che i risultati che otteniamo non siano il valore vero della grandezza, bensì costituiscano una misura tanto più precisa quanto più si avvicina al cosiddetto valore vero?
I tipi di errore che possiamo commettere all'atto della misura sono molteplici e si possono racchiudere in tre gruppi fondamentali:

- **Errori casuali**
- **Errori sistematici**
- **Disturbi**

Inoltre si può individuare una quarta categoria, quella rappresentata dai veri e propri **svarioni**.

ERRORI CASUALI

Si dicono ***casuali*** tutti quegli errori che possono avvenire, con la stessa probabilità, sia in difetto che in eccesso.

Data questa caratteristica, definiamo ***errori casuali*** tutte quelle incertezze sperimentali che possono essere rilevate mediante la ripetizione delle misure.
Questi tipi di errore si possono manifestare per svariati motivi: ad esempio a causa della variazione del tempo di reazione da un soggetto ad un altro (e anche per lo stesso soggetto in situazioni diverse), per errori di lettura di indici dovuti ad un non perfetto allineamento tra l'osservatore e la scala graduata o imputabili ad una interpolazione errata, o anche per semplici fluttuazioni del sistema in esame attribuibili per esempio a degli sbalzi termici.

La loro natura di casualità è proprio legata al fatto che essi hanno un'origine aleatoria e molto spesso temporanea: questo, al ripetersi delle misure, determina sull'evento in esame delle fluttuazioni in modo tale che le misurazioni che si ottengono oscillano attorno ad valore pressoché costante.
Ovviamente nel caso in cui sia possibile ripetere le misure l'individuazione di tali errori è

abbastanza semplice: inoltre all'aumentare del numero delle misure, le fluttuazioni introdotte tendono a "bilanciarsi" in quanto avvengono sia in difetto che in eccesso con la stessa probabilità.

ERRORI SISTEMATICI

Le incertezze sperimentali che non possono essere individuate attraverso la ripetizione delle misure sono dette ***sistematiche***.

In particolare si è definito ***errore sistematico*** la differenza tra il valore reale della grandezza in esame e il valore assunto dalla misura effettuata su di essa: ovviamente il valore reale della grandezza in genere non lo si conosce, altrimenti non avrebbe neppure senso effettuare la misura. La principale peculiarità di questo tipo di errori è la loro difficile individuazione e valutazione in quanto, abbiamo detto, non sono riconoscibili attraverso la ripetizione delle misure: è compito quindi di chi esegue le misure accertare la presenza di tali errori in base ad una "sensibilità sperimentale" che si acquisisce soprattutto con la pratica.

Già da questi brevi cenni appare chiaro che il trattamento degli errori sistematici è tanto importante quanto delicato. La loro natura subdola, che "forza" la misura sempre nello stesso verso, ne rende difficile l'individuazione, nonostante ciò si conoscono le fonti principali di tali errori e qui di seguito ci limiteremo ad elencare le più importanti.

- Difetto dello strumento usato
- Interazione strumento-sperimentatore
- Interazione strumento-fenomeno in esame
- Errate condizioni di lavoro
- Imperfetta realizzazione del fenomeno

I metodi per individuare tali errori sono fondamentalmente due, dettati direttamente dalla natura dell'errore sistematico: o si ricorre ad una previsione teorica del fenomeno e la si confronta con le misure ottenute, oppure si eseguono ulteriori misure utilizzando apparati diversi che evidenzino la presenza (ma non sempre la natura) di tali errori.
Mentre attraverso il confronto con una previsione teorica può rimanere il dubbio che quest'ultima sia errata, ripetendo le misure con strumenti diversi, l'individuazione di possibili errori sistematici attraverso le discrepanze tra i valori ottenuti risulta molto più agevole.

DISTURBI

A volte può accadere che l'accuratezza di una misura si riduca a causa di alcuni fenomeni detti ***disturbi***. Con disturbi si intendono tutte le risposte di uno strumento non generate dalla grandezza posta sotto osservazione.

Nella risposta globale dello strumento la presenza di disturbi nel prelievo della misura provoca una sovrapposizione tra il valore effettivo della grandezza e quello fittizio introdotto da questi: la possibilità che il risultato della misura coincide con il valore effettivo si affievolisce fino a svanire, nei casi peggiori.
Ma come si comportano i disturbi nei confronti della misura?

Fondamentalmente i disturbi possono interagire con la misura in due modi:

- Il loro valor medio è circa nullo: allora hanno un comportamento riconducibile a quello di fluttuazioni casuali.
- Il loro valor medio è diverso da zero: in questo caso manifestano una natura di tipo sistematico in quanto il loro effetto va a sommarsi, nella misura, al valore effettivo della grandezza.

Le loro origini possono essere molteplici. Possono derivare dallo strumento di misura, dal modo in cui viene usato o dall'ambiente in cui viene adoperato. Possono però anche essere imputabili all'interazione con altre grandezze che fanno parte del fenomeno che stiamo osservando, ma che non sono oggetto della nostra misura: ad esempio, un fenomeno abbastanza diffuso con il quale ci si deve spesso confrontare è quello dell'agitazione termica.

- DISTURBI - Altri esempi

Esistono molti esempi di disturbi interagenti con il sistema che pregiudicano le misure che si vogliono prelevare: ne daremo qui alcuni esempi.

I disturbi più classici sono quelli legati all'ambiente nel quale si svolge l'esperimento: un esempio famoso è l'esperimento di Michelson - Morley (1887) per la misurazione della velocità assoluta della Terra "nell'etere" o, come era chiamato allora, il "vento d'etere". Senza entrare nei dettagli dell'esperimento, per capire l'importanza di eliminare ogni tipo di vibrazioni nell'osservazione del fenomeno e nel prelievo delle misure. Si pensi che tutt'intorno all'istituto di fisica di Cleveland nel quale si svolgeva l'esperimento fu fermato il traffico per l'intera durata delle operazioni: le vibrazioni, infatti, avrebbero potuto falsare l'esito e le misurazioni dell'esperimento.

Se osserviamo attentamente con una lente d'ingrandimento un'emulsione fotografica sviluppata senza essere stata precedentemente impressionata, si possono notare dei granuli di argento metallico che costituiscono il cosiddetto fondo.

Il *rumore Johnson* dovuto al moto di agitazione termica delle cariche elettriche di conduzione: questo fenomeno si evidenzia nel momento in cui andiamo a fare una misura abbastanza accurata della differenza di potenziale (ddp) ai capi di un conduttore al quale non è applicata alcuna ddp.
La misura della ddp dovrebbe dare, ovviamente, un risultato nullo, mentre invece si osserva una ddp di ampiezza casuale senza una frequenza fissa con un effetto medio nel tempo nullo.
Queste fluttuazioni attorno al valore zero che ci si aspetterebbe sono dovute all'agitazione termica delle cariche e agli urti tra queste e il reticolo cristallino del conduttore che le contiene.

SVARIONI

I cosiddetti *svarioni* (in inglese *blunders*) racchiudono tutti gli errori madornali: errori sulla lettura dello strumento, sulla trascrizione di dati, errori dovuti a problemi di trasmissione dati in strumenti

digitali o a transienti (variazioni brusche) nell'alimentazione degli apparati di misura.
Meritano di essere considerati a parte poiché non sono del tutto assimilabili né agli errori di tipo casuale, mentre producono effetti di tipo sistematico se sono pochi: se invece, per esempio, si commettono tanti svarioni, questi alla fine possono anche compensarsi.

GLI ERRORI: ASSOLUTI E RELATIVI

Una qualsiasi misura si può esprimere attraverso la scrittura:

$x = x_m \pm \delta x$

L'errore δx, rappresenta l'intervallo entro il quale ragionevolmente riteniamo che la quantità misurata si trovi, ma di per se stesso non ci fornisce un quadro completo della situazione.
Vediamo allora come ampliare la nostra discussione sugli errori introducendo un'importante distinzione: quella tra errori assoluti ed errori relativi.

GLI ERRORI ASSOLUTI

Data la scrittura

$x = x_m \pm \delta x$

definiamo **errore assoluto** di una misura la quantità δx: essa rappresenta l'intervallo entro il quale siamo convinti che il valore vero della grandezza si trovi.

A seconda che il risultato da noi fornito sia frutto di una singola misura o sia il passo finale dell'elaborazione di un gruppo di misure effettuate sulla grandezza in esame, l'errore assoluto dipende quasi totalmente dal metodo di misura adottato e dagli strumenti usati, oppure è correlato a particolari parametri statistici ottenibili dallo studio del gruppo di misure raccolto.
In particolare se abbiamo effettuato una sola misura della grandezza, l'errore assoluto che andremo ad associare alla nostra stima sarà legato a problemi di interpolazione, difficoltà nella lettura della scala dello strumento, sensibilità dello strumento usato o a qualsiasi altra fonte di errore casuale o sistematico.
Se invece, per la valutazione della stima della grandezza, abbiamo a disposizione un gruppo di dati costituiti da tutte le misure ripetute che abbiamo effettuato, possiamo ragionevolmente ritenere che queste si siano distribuite secondo una distribuzione normale (ovviamente in assenza di errori di tipo sistematico) e assumere come errore assoluto da associare alla singola misura la cosiddetta

deviazione standard . Questa è rappresentata dalla radice quadrata della varianza fratto il numero totale (*N*) di misurazioni effettuate, ossia:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N}}$$

In ultimo si noti che sia che venga definito attraverso la deviazione standard o avvalendosi delle caratteristiche degli strumenti usati, l'errore assoluto ha la stessa unità di misura della gandezza osservata.

GLI ERRORI RELATIVI

L'errore assoluto ci da un'indicazione dell'intervallo entro in quale ci aspettiamo ragionevolmente che si trovi il valore vero della grandezza osservata però non descrive un quadro completo della situazione.
Infatti mentre un errore di un centimetro su una distanza di un chilometro indicherebbe una misura insolitamente precisa, un errore di un centimetro su una distanza di quattro centimetri indicherebbe una valutazione piuttosto rozza.
Diventa allora importante considerare l'**errore relativo** (o *errore frazionario*) così definito:

$$errore\ relativo = \frac{\delta x}{x_m}$$

dove δx è l'errore assoluto e $\overline{x}$ è la nostra miglior stima.
Contrariamente all'errore assoluto, l'errore relativo non ha le stesse dimensioni della grandezza, ma è adimensionale: per questo motivo molto spesso si preferisce esprimerlo in forma percentuale, moltiplicandolo per 100. Ad esempio:

grandezza $x = (75 \pm 3)$ unità x

ha un errore relativo pari a:

$$\frac{\delta x}{x_m} = \frac{3}{75} = 0.04$$

di conseguenza un errore in forma percentuale pari al 4 %.

L'errore relativo costituisce un'indicazione della qualità di una misura, in quanto esso rappresenta la precisione della misura stessa.

GLI SCARTI E LO SCARTO QUADRATICO MEDIO

Supponiamo di aver ricavato *N* misure della stessa grandezza *x*: da queste abbiamo poi calcolato la miglior stima attraverso la media e ora ci apprestiamo a dare una valutazione dell'incertezza da associare a tale stima.
Iniziamo col considerare una prima quantità chiamata **scarto** o **deviazione**. Tale grandezza è così definita:

$$d_i = x_i - \overline{x}$$

Questa differenza ci da una indicazione di quanto la i-esima misura $x_i$differisce dalla media: in generale se tutti gli scarti sono molto piccoli allora le nostre misure saranno tutte vicine e quindi presumibilmente molto precise. Al di là del valore numerico degli scarti, sinonimo di precisione o meno nelle misure, è interessante notarne il segno: le deviazioni possono essere infatti sia positive che negative a seconda che l'i-esimo dato cada a destra o a sinistra della media.
Questo ci complica un po' la vita: infatti se volessimo provare a valutare l'incertezza attraverso una media dei singoli scarti ci accorgeremmo subito che la media degli scarti è uguale a zero.
Non dovremmo però rimanere troppo stupefatti davanti a questo risultato in quanto la media, per sua definizione, è tale per cui i dati si distribuiscono sia alla sua sinistra che alla sua destra facendo si che la somma tra gli scarti negativi e quelli positivi sia appunto nulla.

Come ovviare allora a questo inconveniente se riteniamo che gli scarti costituiscano un buon punto di partenza per lo studio dell'incertezza da associare alla media?
Il modo più semplice per salvare capra e cavoli è quello di elevare al quadrato le singole deviazioni ottenendo tutte quantità positive e quindi in grado di essere sommate tra loro senza incorrere in un risultato nullo.
Dopodiché si può passare a calcolare la media estraendone la radice quadrata per ottenere una grandezza compatibile, a livello di unità di misura, con quella di partenza: la grandezza così ottenute è detta **deviazione standard**.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N}}$$

La deviazione standard così ottenuta si rivela molto utile per quantificare l'intervallo entro il quale si distribuiscono le varie misure: in particolare vedremo, attraverso lo studio della distribuzione normale che il 68% delle nostre misure dovrebbe trovarsi all'interno dell'intervallo centrato sulla media e di estremi $+\sigma$ e $-\sigma$. Si può inoltre assumere la deviazione standard come errore da associare al valore medio della misura: così facendo siamo sicuri al 68% di aver individuato l'intervallo entro il quale il valore vero della grandezza dovrebbe cadere.

LA DEVIAZIONE STANDARD DELLA MEDIA

La deviazione standard, calcolata su un gruppo di ***N*** misure, assolve bene il compito di incertezza da associare alla singola misura della grandezza in esame, mentre per quello che riguarda l'incertezza sulla media si ricorre ad un'altra grandezza ancor più idonea allo scopo.
Tale grandezza è la *deviazione standard della media* ed è definita come

$$\hat{\sigma} = \frac{\sigma}{\sqrt{N}}$$

Utilizzando questa nuova grandezza come incertezza da associare alla media di ***N*** misure, abbiamo in realtà fatto un piccolo assunto. Abbiamo cioè supposto che le singole misure effettuate si siano distribuite seguendo la distribuzione di Gauss (assunto peraltro assai fondato) e di conseguenza abbiamo considerato la deviazione standard più come incertezza sulle singole misure che sulla media di quest'ultime.

LA DEVIAZIONE STANDARD DELLA MEDIA
- Dimostrazione -

Supponiamo di aver realizzato *N* misure della stessa grandezza e che i singoli valori ottenuti x_i si siano distribuiti normalmente attorno al valor medio *X* con larghezza σ: vogliamo a questo punto conoscere l'affidabilità della "media delle *N* misure".
Per fare ciò immaginiamo di ripetere il nostro gruppo di *N* misure molte volte e per ciascuna di esse andiamo a calcolare la media. Quello che vogliamo fare è trovare la distribuzione di queste determinazioni della media delle *N* misure.
Calcoliamo, in ciascuna sessione di misure, la media

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_N}{N}$$

Essendo il valore medio di ciascuna delle x_i pari a *X*, il valore di $\overline{x}$ risulta

$$\overline{x} = \frac{X + X + \ldots + X}{N} = X$$

E fin qui tutto bene, nel senso che dopo aver effettuato molte determinazioni della media $\overline{x}$ di *N* misure, i nostri molti risultati per $\overline{x}$ si saranno distribuiti normalmente attorno al valore vero *X*.
L'unico problema che ci resta da risolvere (ed è quello che ci interessa) è quello di trovare la larghezza della nostra distribuzione di risultati. Essendo fondamentalmente la media funzione di tutte le *N* misure x_i, la la formula generale per la propagazione delle incertezze per funzioni di più variabili ci da

$$\hat{\sigma} = \sqrt{\left(\frac{\partial \overline{x}}{\partial x_1}\sigma_{x_1}\right)^2 + \ldots + \left(\frac{\partial \overline{x}}{\partial x_N}\sigma_{x_N}\right)^2}$$

Dal momento che le misure x_i sono tutte misure della stessa grandezza, le loro larghezze sono tutte le stesse ed uguali a σ; inoltre tutte le derivate parziali presenti nella formula sono uguali ad 1/N data la definizione stessa di media scritta poc'anzi.
Allora abbiamo

$$\hat{\sigma} = \sqrt{\left(\frac{1}{N}\sigma\right)^2 + \ldots + \left(\frac{1}{N}\sigma\right)^2}$$

ossia otteniamo

$$\hat{\sigma} = \frac{\sigma}{\sqrt{N}}$$

LA PROPAGAZIONE DEGLI ERRORI

La maggior parte delle quantità fisiche non può essere misurata attraverso una singola misura diretta, ma occorre determinarla attraverso due passi distinti: la misura diretta delle singole grandezze e, attraverso queste, il calcolo della quantità cercata.
Per esempio, per misurare la superficie di un tavolo rettangolare, occorre prima effettuare direttamente le misure dei due lati valutando le relative incertezze, dopodiché si passa a calcolare la superficie attraverso il prodotto dei due lati.
Ora, come l'operazione di misura comporta due passi, anche la determinazione dell'errore necessita di due fasi: dapprima occorre valutare le incertezze delle grandezze misurate direttamente, e di conseguenza si deve trovare come tali incertezze si propaghino attraverso i calcoli

Studiando le operazioni più semplici tra le grandezze misurate direttamente dovremo analizzare i seguenti casi:

Sfruttando le regole ricavate dall'analisi di questi casi potremo passare a studiare anche funzioni composte riconducibili a singole applicazioni delle operazioni sopraelencate attuando lo studio della cosiddetta propagazione passo per passo.

Nel caso in cui la grandezza da calcolare sia una funzione più complicata delle quantità iniziali, si ricorrerà alla regola generale per il calcolo dell'errore di una funzione di più variabili.

ERRORI NELLE SOMME E NELLE DIFFERENZE

Per vedere come si propagano gli errori nelle somme e nelle differenze consideriamo il seguente esempio.
Supponiamo di avere le misure dirette (*x* e *y*) delle due grandezze di cui si vuol calcolare la somma così espresse:

$x = x_m \pm \delta x$

$y = y_m \pm \delta y$

Se chiamiamo *z* la grandezza pari alla somma di *x* e *y*, abbiamo che il valore "più alto" probabile per *z* si ha quando *x* e *y* assumono i loro valori più grandi, cioè per $x_m + \delta x$ e $y_m + \delta y$

valore massimo probabile = $(x_m + \delta x + y_m + \delta y) = (x_m + y_m) + (\delta x + \delta y)$

mentre il valore "più basso" probabile si ha quando entrambe *x* e *y* assumono il loro minino ossia $x_m - \delta x$ e $y_m - \delta y$, per cui si ha

valore minimo probabile = $(x_m - \delta x + y_m - \delta y) = (x_m + y_m) - (\delta x + \delta y)$

osservando i due valori (massimo e minimo) ricavati si vede che per la grandezza derivata $z=x+y=$ $z_m \pm \delta z$ abbiamo

$z_m = x_m + y_m$

e

$\delta z \approx \delta x + \delta y$

Si può operare in modo analogo per calcolare l'errore commesso nel caso di una differenza e si raggiunge lo stesso risultato. Possiamo allora definire la regola della composizione degli errori in una somma o in una differenza, generalizzando il risultato ottenuto a somme o differenze di *N* termini, nel modo seguente:

> *se diverse grandezze x, y, ... , w sono misurate con incertezze* δx, δy, ... , δw *e tali valori vengono utilizzati per calcolare quantità del tipo*
>
> $$z=x+...+y-(u+...+w)$$
>
> *allora l'errore nel valore calcolato di z è pari alla somma*
>
> $$\delta z \approx \delta x + ... + \delta y + \delta u + ... + \delta w$$
>
> *di tutti gli errori assoluti originali.*

Si è usato il simbolo di uguaglianza approssimata ($\approx$) per sottolineare il fatto che quella che è stata fornita è una valutazione un po' "grezza" dell'errore δz. Vedremo più avanti che esiste una stima migliore costituita dalla somma quadratica (o somma in quadratura) dei singoli errori.
Per ora accontentiamoci della definizione data con la consapevolezza che essa rappresenta un **limite superiore** per la valutazione dell'errore valido in qualsiasi caso.

ERRORI NEI PRODOTTI E NEI QUOZIENTI

Supponiamo che la grandezza che vogliamo calcolare sia $z = xy$: nel calcolo dell'errore nei prodotti (o nei quozienti) sfruttiamo gli errori relativi delle singole misure.
Utilizziamo però una forma leggermente diversa per esprimere i valori delle quantità misurate direttamente in termini di errore relativo, cioè:

valore misurato di $x = x_m \pm \delta x = x_m \left(1 \pm \frac{\delta x}{|x_m|}\right)$

valore misurato di *y*= $y_m \pm \delta y = y_m \left(1 \pm \frac{\delta y}{|y_m|}\right)$ Ora, poiché x_m e y_m costituiscono le nostre migliori stime per le grandezze *x* e *y*, la stima migliore che possiamo fare della grandezza *z* è proprio

$$z_m = x_m \cdot y_m$$

Analogamente al caso della somma e della differenza, il valore massimo probabile per *z* si ha quando vengono considerati i valori più grandi di *x* e *y*, ossia nel caso in cui abbiamo il segno più nell'errore relativo:

valore massimo probabile = $x_m \cdot y_m \left(1 + \frac{\delta x}{|x_m|}\right)\left(1 + \frac{\delta y}{|y_m|}\right)$ mentre il valore minimo è dato da

valore minimo probabile = $x_m \cdot y_m \left(1 - \frac{\delta x}{|x_m|}\right)\left(1 - \frac{\delta y}{|y_m|}\right)$ se ora andiamo a sviluppare il prodotto delle due parentesi nel caso del valore massimo (il discorso si può ripetere in modo analogo per il valore minimo) otteniamo:

$$\left(1 + \frac{\delta x}{|x_m|}\right)\left(1 + \frac{\delta y}{|y_m|}\right) = 1 + \frac{\delta x}{|x_m|} + \frac{\delta y}{|y_m|} + \frac{\delta x}{|x_m|}\frac{\delta y}{|y_m|} \simeq 1 + \frac{\delta x}{|x_m|} + \frac{\delta y}{|y_m|}$$

Si noti che nell'ultimo passaggio si è usata l'uguaglianza approssimata poiché abbiamo trascurato il prodotto dei due errori relativi in quanto, essendo singolarmente numeri piccoli (dell'ordine di qualche percento), il loro prodotto è assai piccolo e trascurabile rispetto altre quantità.

Ne concludiamo in definitiva che le misure siffatte di *x* e *y* portano ad un valore di *z=xy* dato da:

$$z = x_m \cdot y_m \left[1 \pm \left(\frac{\delta x}{|x_m|} + \frac{\delta y}{|y_m|}\right)\right]$$

Dal confronto di questa con la forma generale per *z*

$$z = z_m \left(1 \pm \frac{\delta z}{|z_m|}\right)$$

possiamo esprimere l'errore relativo su *z* come la somma degli errori relativi di *x* e *y*

$$\frac{\delta z}{|z_m|} \approx \frac{\delta x}{|x_m|} + \frac{\delta y}{|y_m|}$$

Per quanto riguarda i quozienti il modo di operare è identico, per cui possiamo enunciare la regola generale della propagazione delle incertezze nei prodotti e nei quozienti nel modo seguente:

se diverse grandezze x, y, ... , w sono misurate con incertezze δx, δy, ... , δw *e tali valori vengono utilizzati per calcolare quantità del tipo*

$$z = \frac{x \ldots y}{u \ldots w}$$ *allora l'errore nel valore calcolato di z è pari alla somma*

$$\frac{\delta z}{|z|} \approx \frac{\delta x}{|x|} + \ldots + \frac{\delta y}{|y|} + \frac{\delta u}{|u|} + \ldots + \frac{\delta w}{|w|}$$

dei singoli errori relativi.
Anche qua abbiamo usato il simbolo di uguaglianza approssimata sempre per sottolineare il fatto che quella che abbiamo ricavato è una sovrastima dell'errore, un **limite superiore**: vedremo come affinare la nostra valutazione dell'incertezza sui prodotti attraverso la somma quadratica (o somma in quadratura) dei singoli errori.

ERRORI NELL'ELEVAMENTO A POTENZA

Per quanto riguarda il calcolo dell'incertezza da associare ad una operazione di elevamento a potenza ci limitiamo a dare la regola generale, rimandando una trattazione più specifica alla sezione dedicata alla regola generale per il calcolo dell'errore in funzione di una o più variabili.

Se x è misurato con incertezza δx *e si deve calcolare l'espressione*

$$z = x^n$$

dove n è un qualunque numero noto fissato, positivo o negativo, allora l'errore relativo di z è |n| volte quello di x, ossia

$$\frac{\delta z}{|z|} = |n| \frac{\delta x}{|x|}$$

ERRORI NEL PRODOTTO CON UNA COSTANTE

Supponiamo di misurare una grandezza *x* e in seguito di utilizzare tale quantità per calcolare il prodotto *z=Kx* dove il numero *K* è una costante e come tale **non ha errore**. Classico esempio di tale situazione è rappresentato dal calcolo della lunghezza di una circonferenza, ove il diametro, misurato con la sua incertezza, viene moltiplicato per la costante pi greca.

Per la valutazione dell'incertezza sul prodotto di una grandezza per una costante ci rifacciamo a quanto è stato detto per il calcolo dell'errore nei prodotti e nei quozienti. In particolare l'errore relativo su <I<Z< i>dovrebbe essere stimabile attraverso la somma di quelli su *K* e su *x*.
Dal momento però che non abbiamo errore su *K* risulta che, anche considerando la somma in quadratura:

$$\frac{\delta z}{|z_m|} = \frac{\delta x}{|x_m|}$$

Se vogliamo considerare l'errore assoluto, basta che moltiplichiamo per $|z_m|=|Kx|$ e otteniamo

$$\delta z = |K|\,\delta x$$

Riassumendo possiamo enunciare la regola per il calcolo dell'errore del prodotto di una grandezza con una costante nel modo seguente:

> *data una grandezza x misurata con la sua incertezza* δx *ed utilizzata per calcolare il prodotto z = Kx , dove K* ***non ha errore****, allora l'errore relativo in z è pari a*
>
> $$\frac{\delta z}{|z_m|} = \frac{\delta x}{|x_m|}$$
>
> *mentre l'errore assoluto è dato da*
>
> $$\delta z = |K|\,\delta x$$

LA PROPAGAZIONE PASSO PER PASSO

Consideriamo ad esempio un'espressione del tipo

$z = x + y\,(u - \ln w)$

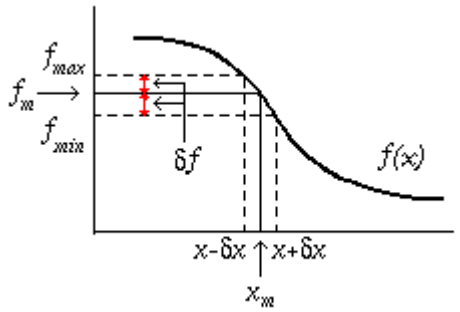
Se partiamo dalle singole quantità misurate *x*, *y*, *u* e *w* possiamo calcolare l'incertezza sulla precedente espressione procedendo in questo modo: calcoliamo la **funzione** *ln w*, quindi la **differenza** *u - ln w*, il **prodotto** *y* (*u - ln w*) e infine la **somma** di *x* con *y* (*u - ln w*).
Dalla precedente analisi sulle singole operazioni siamo in grado di dire come si propagano gli errori attraverso ogni singolo passaggio, cosicché, se supponiamo come abbiamo fatto fin'ora che le grandezze in esame siano indipendenti, calcoliamo l'errore sul risultato finale procedendo secondo i passi descritti partendo dalle misure originali.
Troveremo così inizialmente l'errore sulla funzione *ln w*: noto questo calcoleremo l'errore sulla differenza *u - ln w* e poi quello sul prodotto *y* (*u - ln w*) arrivando finalmente all'errore completo sull'espressione *x + y* (*u - ln w*).

Bisogna però prestare attenzione al fatto che trattando allo stesso tempo somme e prodotti, si deve possedere una certa dimestichezza nel trattare simultaneamente errori assoluti ed errori relativi: inoltre abbiamo parlato della funzione logaritmo, ma non è stato ancora illustrato il metodo per ricavare l'incertezza associata a tale operazione.
Per questa e in generale per funzioni arbitrarie di una variabile si rimanda alla sezione dedicata allo studio dell'incertezza per funzioni di una variabile o alla generalizzazione a più variabili.

- PROPAGAZIONE DEGLI ERRORI - FUNZIONI ARBITRARIE DI UNA VARIABILE

Dopo avere studiato i casi di somma, differenza, prodotto e quoziente andiamo a studiare funzioni più complicate di una variabile e cerchiamo di trovare una regola generale per la propagazione degli errori in tali funzioni.

Vediamo come procedere: supponiamo al solito di aver misurato una grandezza *x* nella forma $x_m \pm \delta x$ e di usare questa quantità per calcolare una qualche funzione nota *f*(*x*).
Per capire come l'errore sulla quantità di partenza *x* si propaghi attraverso il calcolo dell'ipotetica funzione *f*(*x*) pensiamo al grafico di quest'ultima: dalla figura vediamo come la miglior stima per *f*(*x*) sia costituita da f_m che non è altro che il valore assunto dalla funzione nel punto x_m. Per quanto riguarda l'errore fruttiamo il più grande e il più piccolo valore probabile di *x*: da questi graficamente troviamo i corrispettivi valori probabili f_{max} e f_{min} della funzione.



Operando in questo modo non è sempre detto che f_{max} e f_{min} siano simmetrici rispetto a f_m: se però l'incertezza δx è sufficientemente piccola la porzione di grafico che andiamo ad analizzare è così ristretta che la funzione in quel dominio può essere approssimata ad una retta. Se così è allora f_{max} e f_{min} sono ugualmente spaziati su entrambe i lati di f_m e l'incertezza δf, che ci permette di scrivere il risultato nella forma $f_m \pm \delta f$, può essere ricavata dal grafico.

Molto spesso però non si ricorre all'osservazione del grafico per ricavare l'errore in quanto si conosce la forma analitica della funzione (ad es. *ln x*, *cos x*, ecc.).
Una nota proprietà dell'analisi matematica afferma che nel caso in cui δy sia piccolo si ha

$$q(y + \delta y) - q(y) = \frac{dq}{dy}\delta y$$

Ora noi abbiamo che se il nostro errore δx sulla misura è piccolo possiamo scrivere

δf = *f*($x_m + \delta x$) - *f*(x_m)

Confrontandola con la precedente possiamo asserire che

$$\delta f = \frac{df}{dx}\delta x$$

La regola generale per il calcolo dell'errore per una funzione arbitraria di una variabile si ottiene da questa con un piccolo accorgimento: poiché la pendenza della curva rappresentante la funzione può

essere sia positiva che negativa, influenzando così il segno della derivata, dobbiamo considerare il valore assoluto di quest'ultima.
In pratica abbiamo:

> *se x, misurato con incertezza δx, viene utilizzato per calcolare una funzione arbitraria f(x), allora l'errore su tale funzione è pari a*

$$\delta f = \left| \frac{df}{dx} \right| \delta x$$

Nel caso in cui la funzione in esame dipenda da più di una variabile, bisogna considerare l'estensione di questa regola al caso di funzioni arbitrarie di più variabili.

- PROPAGAZIONE DEGLI ERRORI - FUNZIONI ARBITRARIE DI PIU` VARIABILI

Introdurremo ora la formula generale per la propagazione degli errori inerente a funzioni di più variabili: attraverso la sua applicazione saremo in grado di risolvere problemi di propagazione delle incertezze.
La necessità di introdurre una formula generale deriva dal fatto che la semplice propagazione passo per passo può talvolta dar luogo a degli errori di valutazione dell'incertezza nel risultato finale: questo accade ad esempio quando in un quoziente la stessa grandezza compare sia al denominatore che al numeratore. In questo caso i due contributi dovuti alle incertezze dei due fattori uguali potrebbero compensarsi parzialmente e attraverso la formula della propagazione passo per passo rischieremmo di dare una sovrastima dell'errore sul rapporto.
Vediamo come procedere partendo da un esempio:

$$z = \frac{x + y}{x + w}$$

Sfruttando il calcolo dell'incertezza per passi, si dovrebbero calcolare gli errori sulle due somme $x + y$ e $x + w$ e successivamente quello sul quoziente rischiando però di trascurare una eventuale interazione tra i due fattori. Se infatti la nostra stima di x fosse errata, ad esempio, in eccesso avremmo che sia il denominatore che il numeratore risulterebbero sovrastimati, ma tale errore di valutazione verrebbe in parte celato se non totalmente cancellato dalla successiva operazione di divisione. Analogamente l'effetto di una eventuale sottostima di entrambe i fattori risulterebbe celato dal quoziente.

Consideriamo allora una funzione qualsiasi di due variabili del tipo

$z = z(x, y)$

Poiché le nostre migliori stime per x e y sono x_m e y_m ci aspettiamo che la migliore valutazione per z sia

$z_m = z(x_m, y_m)$

Per quanto riguarda il calcolo dell'incertezza su *z* dobbiamo ampliare il discorso già fatto nel caso di una variabile: dobbiamo cioè introdurre la seguente approssimazione per una funzione di due variabili

$$z(x+\delta x, y+\delta y) \approx z(x,y) + \frac{\partial z}{\partial x}\delta x + \frac{\partial z}{\partial y}\delta y$$

Tenendo presente che i valori probabili degli estremi per *x* e *y* sono dati da $x_m \pm \delta x$ e $y_m \pm \delta y$ e che le derivate parziali in *x* e *y* possono essere sia positive che negative, otteniamo che i valori estremi di *z* sono dati da

$$z(x_m, y_m) \pm \left(\left|\frac{\partial z}{\partial x}\right| \delta x + \left|\frac{\partial z}{\partial y}\right| \delta y \right)$$

Da questa ricaviamo subito che l'errore su *z* (*x* , *y*) è proprio

$$\delta z \approx \left|\frac{\partial z}{\partial x}\right| \delta x + \left|\frac{\partial z}{\partial y}\right| \delta y$$

Se consideriamo la somma in quadratura possiamo rimpiazzare il risultato ottenuto con una stima più affinata: definiamo allora una regola generale per la propagazione degli errori in funzioni di più variabili.

> *Siano x, y, ... , w misurati con incertezze* ***indipendenti*** *tra loro e* ***casuali*** δx, δy, ... , δw *e tali valori vengano utilizzati per calcolare la funzione z (x , y, ... ,w), allora l'errore su z è dato da*
>
> $$\delta z = \sqrt{\left(\frac{\partial z}{\partial x}\delta x\right)^2 + \left(\frac{\partial z}{\partial y}\delta y\right)^2 + \ldots + \left(\frac{\partial z}{\partial w}\delta w\right)^2}$$
>
> *In ogni caso esso è limitato superiormente dalla somma*
>
> $$\delta z \leq \left|\frac{\partial z}{\partial x}\right| \delta x + \left|\frac{\partial z}{\partial y}\right| \delta y + \ldots + \left|\frac{\partial z}{\partial w}\right| \delta w$$

Da questa regola si possono facilmente ricavare tutte la altre già viste nei casi di operazioni più semplici, mentre l'uso diretto della formula può risultare un po' macchinoso: quando è possibile si preferisce procedere con il calcolo passo per passo ricordando però che nel caso in cui la stessa variabile compaia più volte nell'espressione questo metodo può portare a delle stime errate.

ANALISI STATISTICA DEI DATI

L'ANALISI STATISTICA
- Introduzione -

Contrariamente a quanto saremmo portati a pensare, la possibilità di attingere ad una massa sterminata di informazioni rischia di impedirci di fatto di utilizzarne anche solo una parte: non basta infatti avere solo l'accesso *teorico* ad una informazione, ma occorre che essa sia effettivamente fruibile.
E` forse questo il problema centrale della **statistica**: rendere davvero utilizzabili grandi quantità di informazioni, teoricamente disponibili, ma di fatto difficilmente gestibili, relative agli oggetti della propria indagine. Infatti tutte le informazioni - per contribuire effettivamente ad accrescere la conoscenza di un fenomeno - hanno bisogno di essere *trattate* da vari punti di vista: occorrono tecniche accurate di rilevazione, occorre procedere a accurate **selezioni**, occorre un lavoro di **organizzazione** e di **sintesi**. D'altra parte il lavoro statistico ha senso solo se si confronta con grandi quantità di informazioni.

Ricapitolando la statistica raccoglie e restituisce in forma organizzata grandi quantità di informazioni. Nel fare ciò "obbedisce" ad una duplice esigenza: quella *predittiva* e quella *descrittiva*.

Ogni comunità sente il bisogno - a fini di documentazione - di raccogliere una serie di dati sugli usi, sui costumi, sulle attività sociali e economiche dei suoi componenti; i censimenti costituiscono uno strumento fondamentale attraverso cui la statistica esplica questa funzione. L'immagine della complessità sociale che ne risulta è parziale e selettiva, ma proprio per questo ha una sua efficacia.
Oltre al carattere descrittivo, un'esigenza, forse la principale, a cui risponde la statistica è quella *predittiva*: la raccolta e l'elaborazione dei dati, e quindi la "fotografia" del passato e del presente, serve per prevedere i comportamenti futuri, per operare scelte, per assumere decisioni.
La statistica, mettendo i dati raccolti ed elaborati a disposizione delle attività di previsione, fornisce i presupposti conoscitivi per orientarsi secondo criteri ragionevoli (anche se non privi di un margine di indeterminazione) nelle situazioni in cui la quantità di informazioni effettivamente utilizzabili si rivela insufficiente a garantire sicurezze.

Si presti infine attenzione al fatto che, essendo l'elaborazione statistica frutto di varie operazioni (selezione, organizzazione e sintesi) può capitare che errori commessi casualmente (o volutamente) in una fase qualsiasi dell'interpretazione dei dati contribuiscano a fornirci un quadro, in parte o completamente, distorto del fenomeno analizzato.

TRAPPOLE STATISTICHE

Con la dicitura *trappole statistiche* intendiamo tutte le fonti di errore o, nel caso in cui vengano sfruttate consciamente, di inganno nella interpretazione dei dati.

Infatti, poiché sappiamo che la statistica serve ad interpretare una realtà fatta di elevatissimi numeri di elementi, abbiamo deciso di fissare la nostra attenzione sulla difficoltà "percettiva" dei numeri, sulle variazioni dovute a trasformazioni numeriche o cambiamenti di scala e sugli effetti di diversi tipi di campionamento.

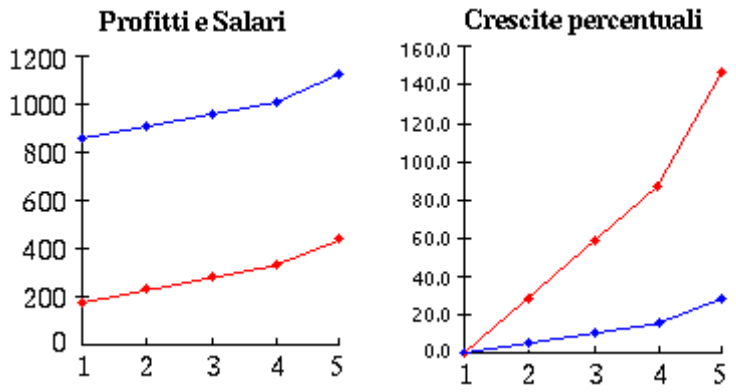
Consideriamo, ad esempio, le trasformazioni numeriche dei dati di partenza. Esse si riflettono molto spesso in un mascheramento se non talora in una distorsione della realtà: si pensi ad esempio a come cambia l'aspetto di un grafico a seconda che si usi una scala lineare o logaritmica oppure si grafichi una nuova variabile traslata rispetto all'origine. Inoltre lo zoom su una porzione della figura può amplificare o ridurre gli effetti interpretativi del fenomeno descritto.
Un'altra trasformazione numerica che può creare problemi è la trasposizione dei dati iniziali in forma percentuale: vediamo a tal proposito un esempio.

Le seguenti tabelle mostrano l'andamento di ipotetici salari e profitti di un'azienda nel corso di cinque anni e i rispettivi incrementi in forma percentuale, riferiti al dato del primo anno.

Anni	Profitti	Salari
1	170	850
2	220	900
3	270	950
4	320	1000
5	420	1100

Anni	+% Profitti	+% Salari
1	0.0	0.0
2	29.4	5.9
3	58.8	11.8
4	88.2	17.6
5	147.1	29.4

La "diversità" tra i due andamenti (la crescita salari - profitti e la crescita delle relative percentuali) si evidenzia ancora meglio nel momento in cui andiamo a rappresentarli graficamente: in blu abbiamo disegnato i salari e la loro crescita percentuale, mentre in rosso abbiamo graficato i dati relativi ai profitti.



Si nota facilmente la differenza tra i due tipi di andamenti.....
Questo piccolo "trucchetto" viene molto spesso usato dai mezzi di informazione per dare una diversa connotazione alla medesima notizia.

MEDIA, MODA, MEDIANA E MOMENTI

LA MEDIA

Quando della stessa grandezza si possiede un campione di valori frutto di *N* misurazioni si possono "riassumere" le informazioni inerenti alla grandezza derivanti dalle singole misure attraverso la media che, in questo caso, costituisce la miglior stima possibile per la grandezza in esame.

Se le singole misure si possono considerare equivalenti l'una all'altra senza che ve ne siano di alcune più importanti o privilegiate, allora definiamo la media aritmetica come segue:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{\sum x_i}{N}$$

Nel caso in cui i singoli dati abbiano pesi diversi, allora si ricorre a quella che viene definita media pesata. Esistono anche altre definizioni di media.

LA MODA

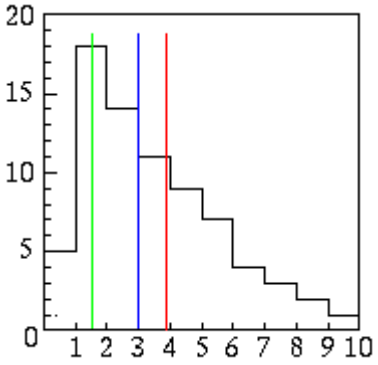
Per quanto riguarda una variabile aleatoria, si definisce *moda* il valore più probabile che questa può assumere: quando trattiamo campioni di dati frutto ad esempio di diverse misure della stessa grandezza, allora definiamo la *moda* come il valore più "popolare" del campione.
Per capire meglio questa definizione pensiamo ad un istogramma: la moda è costituita in questo caso dal valore corrispondente alla colonna più alta.

LA MEDIANA

La mediana è, ad esempio in un istogramma, l'ascissa corrispondente al punto in cui l'area delimitata dall'istogramma si divide in due parti uguali: in pratica il numero di dati che sta alla destra della mediana (quelli maggiori) è uguale al numero di dati alla sinistra della mediana (quelli minori).

Nota

LA MEDIA, LA MODA E LA MEDIANA



Non sempre queste tre grandezze coincidono: anzi solo in rari casi in cui la nostra distribuzione di dati o il nostro istogramma godono di particolari simmetrie può avvenire che queste coincidano.
A dimostrazione di questo si veda l'istogramma a lato.
Nonostante ciò è utile notare come il fatto che i dati siano ammassati verso i primi valori dell'istogramma, dimostrando una certa asimmetria, faccia sì che le tre quantità (*media* in **rosso**, *mediana* in **blu** e *moda* in **verde**) non siano coincidenti.
In particolare in questo istogramma si sono considerati 10 intervalli, corrispondenti ai valori da 1 a 10 in ascissa e un totale di 74 dati. La media è pari a 3.94, la mediana ha valore 3 e la moda è rappresentata dall'intervallo in ascissa 1-2.

I MOMENTI

Accanto alle cosiddette *caratteristiche di posizione* quali la media, la moda e la mediana, esistono altre caratteristiche, ciascuna delle quali descrive una proprietà della distribuzione in esame.
I **momenti** forniscono indicazioni sulla dispersione rispetto al valore centrale, sull'asimmetria della distribuzione, ecc.: nelle applicazioni pratiche si ha a che fare con *momenti iniziali* e con *momenti centrali.*

LA SCELTA DEI DATI

Quando si hanno a disposizione diverse misure della stessa grandezza, può accadere che una o più misure siano in disaccordo con le restanti: quando ciò si verifica entra in gioco in modo preponderante la figura dello sperimentatore. E` lui infatti che dovrà decidere il dato è compatibile o meno con gli altri e di conseguenza se conservarlo o *rigettarlo*.
Il fatto che in ultima analisi sia lo sperimentatore a decidere delle sorti del dato fa sì che il "criterio" per il rigetto o meno dei dati sia, alla fine, soggettivo e quindi diventi quasi impossibile delinearne un quadro preciso.
In verità esistono dei criteri già formalizzati e accettati dalla maggior parte degli scienziati: l'influenza dello sperimentatore, al momento dell'utilizzo di questi metodi risiede nello stabilire quale sia la "soglia di accettabilità" del dato.
Abbiamo detto che esistono dei criteri già formulati e, nella maggior parte dei casi, accettati tra cui citiamo:

- Il criterio di Chauvenet
- Il criterio *"a priori"*

E inoltre, ma merita un discorso a parte:

- Il test χ^2

Anche se attraverso metodi e procedimenti differenti, in linea di principio questi criteri forniscono a chi li usa una stima di quanto un dato si discosta dagli altri: è qui che interviene lo sperimentatore definendo la soglia oltre la quale un dato deve essere rigettato o meno. Tale soglia dipende da diversi fattori tra cui la metodologia e le tecniche di misura adottate.

La determinazione del valore oltre il quale scartare può rappresentare un'arma a doppio taglio per due motivi fondamentali.
Come prima cosa uno scienziato può essere tacciato di voler pilotare i dati del suo esperimento, scartando quelli che non sono in accordo con le sue previsioni teoriche e in secondo luogo esiste il rischio di scartare eventi considerandoli frutto di qualche errore, quando questi potrebbero essere il segnale di qualche importante fenomeno che lo scienziato aveva trascurato o del quale non era a conoscenza.
A questo proposito va sottolineato che molte importanti scoperte sono state frutto di misure all'apparenza anomale. Tali misure, prima di essere scartate, sono state analizzate in modo più approfondito: si è quindi visto che, quello che sembrava un evento scarsamente significativo, rappresentava invece un segnale di una realtà diversa da come la si era formalizzata.

LA SCELTA DEI DATI
- Il criterio di Chauvenet -

Per capire il criterio di Chauvenet per il rigetto dei dati consideriamo un esempio. Supponiamo di aver effettuato dieci misure di una certa grandezza *X* e di averle riassunte nella seguente tabella:

1	2	3	4	5	6	7	8
14.1	13.4	13.8	13.0	11.8	14.1	13.0	14.0

Se ora procediamo al calcolo della media ($\overline{x}$) e della deviazione standard (σ) troviamo i valori:

$\overline{x} = 13.4$

$\sigma = 0.8$

In questa serie di misure il quinto valore (11.8) è decisamente in disaccordo con tutti gli altri: vediamo come procedere nei confronti di tale valore.
Dobbiamo prima di tutto quantificare quanto la misura in questione sia anomala rispetto alle altre: per fare questo, notiamo che il valore 11.8 si discosta dal valor medio di due volte la deviazione standard.
Se assumiamo che le misure si conformino ad una distribuzione di Gauss avente media $\overline{x}$ e deviazione standard σ allora siamo in grado di calcolare la probabilità di avere misure che differiscano dalla media di almeno due deviazioni standard.
La probabilità di avere tali misure si ottiene, secondo la proprietà degli eventi contrari, sottraendo

da uno (il 100% rappresentante la globalità degli eventi) la probabilità di ottenere risultati *entro* due deviazioni standard, cioè:

$$\overline{P_{2\sigma}} = 1 - P_{2\sigma}$$

dove con $\overline{P_{2\sigma}}$ si intende appunto la probabilità di ottenere valori al di fuori di 2σ e con $P_{2\sigma}$ la probabilità di ottenerne entro 2σ.
Da quanto detto, andando a vedere il valore tabulato di $P_{2\sigma}$, otteniamo:

$$\overline{P_{2\sigma}} = 1 - 0.95 = 0.05$$

In pratica abbiamo il 5% di probabilità di ottenere una misura al di fuori di due deviazioni standard, cioè ci si deve aspettare che una misura su venti si discosti di più di 1.6 unità (2σ) dal valor medio (che nel nostro caso era 13.4).

Avendo noi eseguito otto misure, per la proprietà delle probabilità di eventi indipendenti, il numero di misure oltre 2σ è dato da:

$n = 0.05 \times 8 = 0.4$

Significa che mediamente ci si dovrebbe aspettare 2/5 di una misura anomala come il nostro 11.8: in questo modo abbiamo quantificato l'anomalia del valore in questione.
Ora si tratta di stabilire quale sia la "soglia di accettabilità" per i dati dopodiché andiamo a vedere se il dato incriminato deve essere rigettato o meno.
Di solito viene stabilita tale soglia ad 1/2, per cui *se il numero atteso (n) di misure anomale è minore di 1/2, la misura sospetta deve essere rigettata*: da questo discende che il nostro valore 11.8 non è da considerarsi ragionevole e quindi deve essere rigettato.

Una volta capito questo esempio, la generalizzazione del criterio ad un problema con più dati è immediata: si supponga di avere *N* misure ($x_1, x_2, \ldots, x_n$) della stessa grandezza *X*.
Come prima cosa calcoliamo $\overline{x}$ e σ dopodiché osserviamo i dati per vedere se esiste qualche valore sospetto. Nel caso ci sia un dato sospetto (x_s) calcoliamo il numero di deviazioni standard (Z_s) di cui x_s differisce da $\overline{x}$ applicando la formula:

$$Z_s = \frac{x_s - \overline{x}}{\sigma}$$

Fatta questa operazione bisogna andare a vedere quale è la probabilità che una misura differisca da $\overline{x}$ di Z_s volte la deviazione standard: per fare questo bisogna ricorrere ai valori della probabilità in funzione del numero di deviazioni standard che si trovano facilmente tabulati.
Alla fine, per ottenere il numero (*n*) di misure anomale che ci si aspetta, moltiplichiamo la suddetta probabilità per il numero totale di misure (*N*):

$n = N \times P$ (oltre $Z_s \sigma$)

Se il numero *n* è minore di 1/2 allora non si attiene al criterio di Chauvenet e come tale deve essere rigettato.

A questo punto si presenta uno spinoso problema:
come agire con i dati rimasti ?
C'è chi sostiene che si debba applicare nuovamente il criterio di Chauvenet ai dati rimasti (tenendo conto che dopo l'eliminazione del primo dato si hanno diversi valori di $\overline{x}$ e σ) fintanto che tutti i dati rimasti siano conformi al criterio di Chauvenet, mentre altri sostengono che tale metodo non vada applicato una seconda volta ricalcolando la media e la deviazione standard. Esiste però anche un terzo modo, forse il più equilibrato anche rispetto a coloro che ritengono che il rigetto di un dato non sia mai giustificato, di affrontare il problema: molti scienziati infatti utilizzano il criterio di Chauvenet non per scartare immediatamente il dato, bensì solamente per *individuarlo*: una volta individuato il valore sospetto si procede alla verifica della sua attendibilità attraverso la riproduzione delle misure e una successiva rianalisi dei dati.

IL CRITERIO DI CHAUVENET
- Note -

È utile sottolineare ulteriormente alcuni concetti di importanza basilare.

-
- In una analisi permane sempre una certa dose di arbitrarietà per quanto riguarda le decisioni da prendere riguardo i dati: tali decisioni possono poi rivelarsi cruciali dal punto di vista della formulazione delle leggi che governano il fenomeno in esame.
- Il procedimento più corretto da seguire nel caso di anomalie nei dati è quello di ripetere la misura molte volte: in questo modo, se l'anomalia persiste, dovrebbe risultare più facile trovarne la causa e quindi eliminarla. Inoltre aumentando sensibilmente il numero di misure l'influenza sul risultato finale dell'eventuale dato sospetto diminuisce in modo significativo.
  Il rovescio della medaglia è costituito dal fatto che, molto spesso è assai scomodo se non addirittura impossibile, ripetere la stessa misura centinaia di volte ogni volta che si presenta un dato anomalo.
  Per questo i criteri studiati per lo scarto o meno dei dati costituiscono un ottimo compromesso per non perdere troppo tempo nella ripetizione delle misure e allo stesso tempo per evitare di scartare prematuramente dei dati che potrebbero rivelare aspetti ignoti del fenomeno.
- Nell'applicazione del criterio di Chauvenet si è supposto di trattare eventi indipendenti cioè che il verificarsi di ciascuno non venisse influenzato dal fatto che gli altri si fossero o meno verificati.

IL CRITERIO "A PRIORI"

Al solito supponiamo di avere una serie di *N* dati di cui calcoliamo la media ($\overline{x}$) e la deviazione standard (σ): in base al criterio a priori il rigetto o meno dei dati viene eseguito rispetto ad una soglia di accettabilità stabilita precedentemente.
In questo modo si fissa un intervallo, eventualmente molto ampio, entro il quale i dati vengono accettati, mentre quelli che cadono al di fuori vengono scartati. Ad esempio, nel caso di una

distribuzione gaussiana, un intervallo relativo alla probabilità dell'1 ‰ ha una semiampiezza di 3.29 σ. La larghezza dell'intervallo va anche considerata in base al numero di misure che si effettuano: se si ha un numero di misure molto inferiori a 1000 si può ritenere che il fatto che un dato sia eventualmente esterno all'intervallo non sia dovuto alla sola fluttuazione statistica, mentre se si è in presenza di qualche migliaio di dati ci si deve aspettare per ragioni statistiche che un certo numero di valori capiti al di fuori dell'intervallo.

IL CRITERIO "A PRIORI"
- Note -

Notiamo che dire che un dato è da scartare, significa che, denominato x_s il valore anomalo, vale la seguente relazione:

$$|x_s - \bar{x}| > Z_s \sigma = Z_s \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}}$$

dove Z_s rappresenta il numero di deviazioni standard che assumiamo come larghezza dell'intervallo. Elevando al quadrato si ha:

$$(x_s - \bar{x})^2 > \frac{Z_s^2}{N-1}(x_s - \bar{x})^2 + \frac{Z_s^2}{N-1}\sum_{i \neq s}(x_i - \bar{x})^2$$

Quest'ultima ci da un limite inferiore ad *N* in quanto si vede che, affinché sia lecito parlare di possibile rigetto del dato, si deve avere $N - 1 > Z_s^2$.

LA SCELTA DEI DATI
- Il criterio del x^2 -

Il test del χ^2 (o, come talvolta viene chiamato, di Pearson) viene usato come una sorta di "verifica delle ipotesi".
Le ipotesi in questione possono riguardare semplicemente la presenza o meno di una correlazione tra diverse variabili (in questo caso si parla di *verifica dell'indipendenza*); oppure riferirsi alla distribuzione teorico-matematica che meglio riproduce i dati sperimentali (allora si parla di *verifica dell'aggiustamento*).
In entrambe i casi il problema è quello di paragonare i risultati sperimentali con le previsioni teoriche e di valutare la distanza globale tra i due insiemi sommando i contributi di ciascun elemento. Questi contributi sono valutati percentualmente, cioè come rapporto tra lo scarto (*differenza tra risultato e previsione*) e la previsione. Gli scarti sono successivamente elevati al quadrato per sopprimere il segno.
La distanza globale che cerchiamo, detta appunto χ^2, è data da:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - T_i)^2}{T_i}$$

dove O_i sono i risultati osservati e T_i le previsioni. Il χ^2 è tanto più grande quanto meno l'osservazione è compatibile con le ipotesi. Siccome il χ^2 è calcolato utilizzando dati sperimentali è esso stesso una variabile aleatoria che segue, appunto, la distribuzione χ^2 nel caso in cui i dati seguano la distribuzione normale.
Esistono tabelle e grafici che forniscono, in funzione del numero di gradi di libertà dell'insieme, cioè del numero di contributi che si sommano, i livelli di confidenza che corrispondono a diversi valori di χ^2, dicono cioè qual'è la probabilità di ottenere un χ^2 maggiore o uguale a quello osservato se l'ipotesi è vera.

L'uso del χ^2 come verifica dell'aggiustamento è concettualmente simile al procedimento che abbiamo descritto. Si costruisce l'istogramma dei dati e si ipotizza che essi seguano una certa distribuzione teorica. Per verificare l'ipotesi si calcola il χ^2 secondo la stessa definizione data sopra, utilizzando come O_i le altezze delle colonne dell'istogramma e come T_i i corrispondenti valori teorici. Anche in questo caso il χ^2 ottenuto è una variabile aleatoria e bisogna consultare le tabelle per poterne derivare una conclusione quantitativa in termini di probabilità. Per la verifica dell'aggiustamento il numero di gradi di libertà è dato da

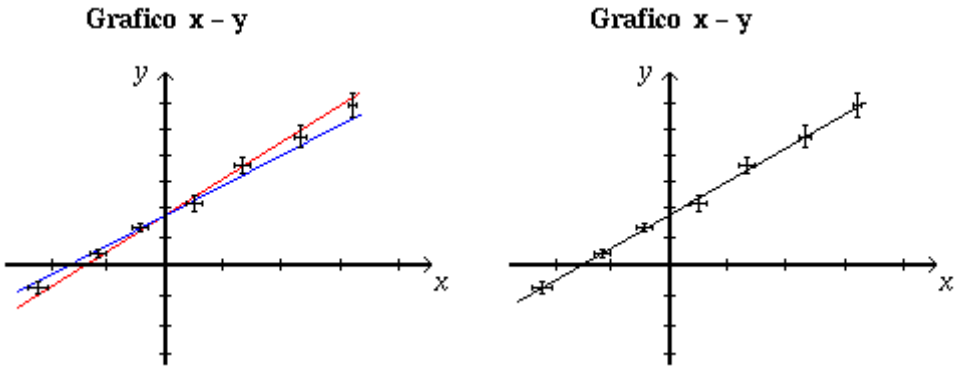
$$n=k-n_p-1$$

dove k è il numero di colonne dell'istogramma e n_p il numero di parametri della distribuzione teorica (ad esempio 3 per la distribuzione di Gauss, 2 per la distribuzione di Poisson, ...).
Se il χ^2 ottenuto è inferiore al valore che secondo le tabelle corrisponde ad un livello di confidenza maggiore o uguale al 95%, di solito l'ipotesi viene accettata: questo significa infatti che, *se i dati seguono effettivamente la distribuzione teorica che stiamo verificando*, la probabilità di avere per fluttuazione statistica un χ^2 minore di quello osservato è soltanto del 5%.

RAPPRESENTAZIONE GRAFICA DEI DATI
- L'interpolazione -

Supponiamo di essere abbastanza soddisfatti della linearità mostrata da un ipotetico insieme di dati: il passo successivo è di chiedersi quale sia la retta che meglio modella i punti sul grafico e successivamente tentare di scoprire quale sia l'equazione che rappresenta la relazione tra le due grandezze.
A causa delle incertezze sperimentali, difficilmente tutti punti di un grafico giacciono su una retta per cui spetta all'abilità dello sperimentatore trovare quale sia la retta che meglio si adatta alla distribuzione dei punti, ossia quella retta che meglio **interpola** i punti del grafico.

Tale retta è detta **retta di best fit**. Nel primo grafico sottostante sono state disegnate le *rette limite* , le rette, cioè, aventi la massima e la minima pendenza tra tutte quelle che si adattano a punti del grafico, mentre il secondo rappresenta proprio la retta di best fit.



RAPPRESENTAZIONE GRAFICA DEI DATI
- La regressione lineare -

Occupiamoci ora del metodo analitico per ricavare la miglior linea retta che interpola una serie di punti sperimentali, metodo chiamato *regressione lineare*.
Per semplificare la trattazione, assumeremo d'ora in poi che le incertezze si abbiano solo sulle grandezza in ordinata e che tutte le incertezze siano uguali. Altro assunto (peraltro ragionevole) che faremo è che ogni misura in *y* sia governata dalla distribuzione di Gauss, con lo stesso parametro σ_y per tutte le misure.

Quello che in pratica vogliamo determinare sono le due costanti **A** e **B** che determinano la retta migliore avente la forma

$y = Ax + B$

Se conoscessimo le costanti **A** e **B** allora per ogni singolo $x_i$potremmo calcolare il corrispondente valore vero di y_i come

$y_i = Ax_i + B$

poiché la misura y_i è governata da una distribuzione normale centrata su questo valore vero con parametro σ_y, la probabilità di ottenere il valore osservato y_i è

$$P_{A,B}(y_i) \propto \frac{1}{\sigma_y} e^{-\frac{(y_i - Ax_i - B)^2}{2\sigma_y^2}}$$

dove i pedici **A** e **B** indicano che questa probabilità dipende dai loro valori che sono incogniti. La probabilità di ottenere il nostro insieme completo di valori osservati $y_1 ... y_N$ è

$$P_{A,B}(y_1, ..., y_N) = P_{A,B}(y_1) \cdot ... \cdot P_{A,B}(y_N) \propto \frac{1}{\sigma_y^N} e^{-\frac{\chi^2}{2}}$$

dove l'esponente è dato da

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - Ax_i - B)^2}{\sigma_y^2}$$

Le migliori stime per le costanti si ottengono imponendo che la probabilità sia massima, in quanto gli $\mathbf{y_i}$ cono i valori effettivamente osservati: questo equivale ad imporre che la somma dei quadrati nell'esponente sia minima. Per trovare tali valori differenziamo rispetto ad **A** e **B** e poniamo le derivate uguali a zero:

$$\frac{\partial \chi^2}{\partial A} = \left(-\frac{2}{\sigma_y^2}\right) \sum_{i=1}^{N} x_i(y_i - Ax_i - B) = 0$$

e

$$\frac{\partial \chi^2}{\partial B} = \left(-\frac{2}{\sigma_y^2}\right) \sum_{i=1}^{N} (y_i - Ax_i - B) = 0$$

Queste due equazioni possono essere riscritte come:

$$A \sum x_i + B\, N = \sum y_i$$

e

$$A \sum x_i^2 + B \sum x_i = \sum x_i y_i$$

Tali equazioni, note anche come *equazioni normali*, una volta risolte, danno la stima delle costanti **A** e **B**

$$A = \frac{N(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\Delta}$$

e

$$B = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{\Delta}$$

dove abbiamo introdotto l'abbreviazione

$$\Delta = N(\sum x_i^2) - (\sum x_i)^2$$

Calcolate le miglior stime delle costanti **A** e **B** dai valori misurati ($\mathbf{x_i}$,$\mathbf{y_i}$) è spontaneo chiedersi quali siano le incertezze nelle nostre stime.

RAPPRESENTAZIONE GRAFICA DEI DATI
- Incertezza nelle costanti A e B -

Prima di passare al calcolo vero e proprio delle incertezze sulle stime di **A** e di **B** vediamo di chiarire alcuni punti sull'incertezza σ_y.

Bisogna ricordare che le $y_1 ... y_N$ *non* sono *N* misure della stessa grandezza: in questo modo non possiamo farci un'idea della loro affidabilità solo esaminando lo sparpagliamento dei loro valori. Possiamo però stimare l'incertezza σ_y nel modo seguente.
Partendo dall'assunto che ogni misura y_i sia normalmente distribuita attorno al suo valore vero Ax_i+B, anche le singole *deviazioni* y_i-Ax_i-B sono normalmente distribuite, con lo stesso valore medio 0 e la stessa larghezza σ_i. Questo suggerisce che una buona stima per σ_i dovrebbe essere data dalla somma del quadrato degli scarti nella forma

$$\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - Ax_i - B)^2$$

Tale stima, però, necessita di essere ulteriormente raffinata: questo è dovuto al fatto che non conosciamo i valori veri delle costanti **A** e **B** bensì li rimpiazziamo con le nostre migliori stime. Questo rimpiazzo riduce leggermente il valore precedentemente definito di σ_i: si può dimostrare che è possibile compensare tale riduzione sostituendo il fattore *N* del denominatore con il nuovo fattore ***N-2***, ottenendo così il risultato finale per σ_i

$$\sigma_y^2 = \frac{1}{N-2}\sum_{i=1}^{N}(y_i - Ax_i - B)^2$$

A questo punto possiamo passare al calcolo vero e proprio delle incertezze sulle costanti **A** e **B**. Essendo le stime di **A** e di **B** funzioni ben definite dei valori misurati $y_1 ... y_N$, le incertezze su tali stime si calcolano semplicemente applicando la propagazione degli errori in termini di quelli in $y_1 ... y_N$, per cui si ottiene

$$\sigma_A^2 = \frac{N\sigma_y^2}{\Delta}$$

e

$$\sigma_B^2 = \sigma_y^2\frac{\sum x_i^2}{\Delta}$$

dove al solito si è introdotta l'abbreviazione

$$\Delta = N\left(\sum x_i^2\right) - \left(\sum x_i\right)^2$$

RAPPRESENTAZIONE GRAFICA DEI DATI
- Adattamento ad altre curve col metodo dei minimi quadrati -

Il caso di due variabili che soddisfano una relazione lineare del tipo **y=Ax+B** non è altro che un caso particolare di una vasta classe di problemi che riguardano le curve di adattamento, molti dei quali possono essere risolti in un modo similare. Vediamo ora alcuni di questi problemi.

Adattamento ad una polinomiale

Spesso accade che ci si aspetti che una variabile *y* sia esprimibile come una funzione polinomiale di una seconda variabile *x*, ossia

$$y = A + Bx + Cx^2 + ... + Wx^n$$

Dato un insieme di osservazioni delle due variabili, si può trovare la migliore stima delle **A,B,...,H** con un procedimento analogo a quello con cui abbiamo stimato le costanti **A** e **B** nel caso della relazione lineare. Per semplificare la trattazione supponiamo che la polinomiale sia di fatto una quadratica del tipo

$$y = Ax^2 + Bx + C$$

Al solito supponiamo di avere una serie di misure con incertezze solo sulla variabile dipendente (**y**) e che tali incertezze siano tutte uguali. Allora per ogni x_i, il corrispondente valore vero di y_i è dato dalla funzione quadratica con **A**, **B** e **C** incogniti. Assumendo che le misure degli y_i siano governate da distribuzioni normali, ciascuna centrata sull'appropriato valore vero e tutte con lo stesso σ_i, possiamo calcolare la probabilità di ottenere proprio i valori osservati $y_1...y_N$ come

$$P(y_1, ..., y_N) \propto e^{-\frac{\chi^2}{2}}$$

dove ora

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - Ax_i^2 - Bx_i - C)^2}{\sigma_y^2}$$

La miglior stima per **A**, **B** e **C** è data da quei valori per cui la probabilità è massima, ossia quando l'esponente χ^2 è minimo.
Differenziando χ^2 rispetto a **A**, **B** e **C** e ponendo queste derivate uguali a 0, otteniamo le tre "equazioni normali"

$$A\sum x_i^2 + B\sum x_i + CN = \sum y_i$$

$$A\sum x_i^3 + B\sum x_i^2 + C\sum x_i = \sum x_i y_i$$

$$A\sum x_i^4 + B\sum x_i^3 + C\sum x_i^2 = \sum x_i^2 y_i$$

Per un dato insieme di misure queste equazioni simultanee per **A**, **B** e **C** possono essere risolte per trovare le miglior stime di **A**, **B** e **C**. Con **A**, **B** e **C** calcolati in questo modo l'equazione

$$y = Ax^2 + Bx + C$$

è chiamata *adattamento polinomiale dei minimi quadrati* per le misure date. Sfortunatamente, soprattutto quando il grado delle polinomiali aumenta, non sempre le equazioni normali sono risolvibili oppure sono estremamente difficili, nonostante questo esiste una grande classe di problemi che possono essere risolti.

Funzioni esponenziali

Data l'importanza della funzione esponenziale in fisica, vediamo anche questo caso. Consideriamo la funzione

$$y = Ae^{Bx}$$

Per ricondurci ad una forma a noi più familiare, applichiamo una "linearizzazione" della funzione applicando la funzione logaritmo: in questo modo otteniamo

$$\ln y = \ln A + Bx$$

Questa operazione, anche se non rende lineare *y* nei confronti di x, fa sì che il logaritmo di *y* lo sia. L'utilità di questa nuova relazione è facilmente verificabile. Se riteniamo che *x* e *y* debbano soddisfare l'equazione esponenziale, allora le variabili *x* e *z*=ln(*y*) dovrebbero soddisfare

$$z = \ln A + Bx$$

Se abbiamo una serie di misure (*x*,*y*), allora per ogni $\mathbf{y_i}$ calcoliamo **z=ln(*y*)**. Le coppie (z_i,y_i) dovrebbero giacere su una linea retta: tale retta può essere adattata con il metodo dei minimi quadrati, come abbiamo già visto, per ricavare le stime delle costanti **A** e **B**.

Regressione multipla

Dopo aver parlato di problemi a due variabili, vediamo un esempio con tre variabili (in molti problemi reali più di due variabili che vanno considerate: si pensi ad esempio alla relazione tra pressione volume e temperatura nei gas).

L'esempio più semplice (noi ci limiteremo a questo) è quello in cui una variabile dipende linearmente dalle altre due:

$$z = A + Bx + Cy$$

Se abbiamo una serie di misure (x_i,y_i,z_i) con tutti gli z_i ugualmente incerti e gli x_i e y_i privi di incertezza, possiamo procedere in modo analogo al caso dell'adattamento ad una retta: in questo caso la miglior stima per le costanti è data dalle equazioni normali

$$AN + B\sum x_i + C\sum y_i = \sum z_i$$

$$A\sum x_i + B\sum x_i^2 + C\sum x_i y_i = \sum x_i z_i$$

$$A\sum y_i + B\sum x_i y_i + C\sum y_i^2 = \sum y_i z_i$$

Questo metodo è chiamato *regressione multipla*

TEORIA DELLE PROBABILITA'

CALCOLO DELLE PROBABILITA`
- Definizioni -

In questa pagina elenchiamo alcune definizioni utili per affrontare lo studio del calcolo delle probabilità.

Caratteristiche numeriche delle variabili aleatorie: Valore atteso di una variabile aleatoria

---

GLI EVENTI

Nel calcolo delle probabilità con la parola evento si intende ogni fatto che in seguito ad una prova può accadere oppure no. Qualche esempio:

- l'apparizione di testa quando si lancia una moneta
- l'apparizione di un asso quando si estrae una carta da un mazzo
- la rivelazione di una data particella in un acceleratore

Ad ogni evento è associato un numero reale che è tanto maggiore quanto più è elevata la possibilità che si verifichi l'evento stesso: chiamiamo tale numero probabilità dell'evento.

EVENTI DIPENDENTI ED INDIPENDENTI

Si dice che l'evento A è **dipendente** dall'evento B se la probabilità dell'evento A dipende dal fatto che l'evento B si sia verificato o meno.

Mentre diciamo che l'evento A è **indipendente** dall'evento B se la probabilità del verificarsi dell'evento A non dipende dal fatto che l'evento B si sia verificato o no.

EVENTI CERTI ED IMPOSSIBILI

Definiamo **evento certo** quell'evento che in seguito ad un esperimento deve obbligatoriamente verificarsi. Tale evento costituisce l'unita di misura per la probabilità: si attribuisce, cioè, all'evento certo probabilità uguale all'unità. Di conseguenza tutti gli altri eventi, probabili ma non certi, saranno caratterizzati da probabilità minori all'unità.

L'evento contrario all'evento certo è detto **impossibile**, ossia un evento che non può accadere nella prova in questione. All'evento impossibile è associata una probabilità uguale a zero.

EVENTI MUTUAMENTE ESCLUDENTESI

Si dicono **eventi mutuamente** escludentesi o *incompatibili* quegli eventi aleatori che non possono verificarsi simultaneamente in una data prova. Ad esempio l'apparizione simultanea di testa e di croce nel lancio di una moneta.

EVENTI EQUIPROBABILI

Degli eventi casuali si dicono **equiprobabili in una data prova** se la simmetria dell'esperimento permette di supporre che nessuno di essi sia più probabile di un altro. Ad esempio l'apparizione di una delle sei facce di un dado nel caso in cui questo sia regolare (non truccato).

EVENTI CONTRARI

Si dicono **eventi contrari** due o più eventi mutuamente escludentesi che formano un gruppo completo. Ad esempio i due eventi - passare l'esame, essere bocciati - costituiscono una coppia di eventi contrari.

SOMMA DEGLI EVENTI

Si definisce **somma** di due eventi A e B l'evento C che consiste nel verificarsi dell'evento A o dell'evento B o di entrambe. La probabilità dell'evento C si scrive nel seguente modo:

$P(C) = P(A \cup B) = P(A + B) = P(A \text{ oppure } B)$

Si veda a questo proposito il teorema di addizione delle probabilità
In generale definiamo la **somma** di un numero qualsiasi di eventi l'evento che consiste nel verificarsi di almeno uno di questi.

PRODOTTO DEGLI EVENTI

Si chiama **prodotto** di due eventi A e B l'evento C che consiste nel verificarsi *simultaneo* degli eventi A e B. La probabilità dell'evento C si indica nel seguente modo:

P(C) = P(A∩B) = P(A×B) = P(A e B)

A questo proposito si veda il teorema di moltiplicazione delle probabilità.
In generale definiamo il **prodotto** di un numero qualsiasi di eventi l'evento che consiste nel verificarsi di tutti questi eventi.

PROBABILITA` CONDIZIONATA

La probabilità che accada l'evento A, calcolata a condizione che l'evento B si sia verificato o meno, si dice **probabilità condizionata** e si denota con:

P(A|B)

GRUPPO COMPLETO DI EVENTI

Si dice che eventi casuali formano un **gruppo completo di eventi** se almeno uno di essi deve definitivamente accadere. Ad esempio l'apparizione dei punteggi 1, 2, 3, 4, 5, 6 nel lanciare un dado costituisce un gruppo completo di eventi.

VARIABILI ALEATORIE

Si dicono **variabili aleatorie** quelle grandezze che posso assumere nel corso di una prova un valore sconosciuto a priori. Si distinguono in variabili aleatorie *discrete* e variabili aleatorie *continue*. Le variabili discrete possono assumere solo un insieme di valori numerabile, mentre i valori possibili di quelle continue non possono essere enumerati in anticipo e riempiono "densamente" un intervallo. Esistono anche variabili aleatorie che assumono sia valori continui che valori discreti: tali variabili sono dette *variabili aleatorie miste*.

VALORE ATTESO DI UNA VARIABILE ALEATORIA

Con il termine **valore atteso** o **speranza matematica** di una variabile aleatoria si è soliti indicare la *somma dei prodotti di tutti i possibili valori della variabile per la probabilità di questi ultimi*.
A seconda che la variabile in questione sia discreta o continua, il *valore atteso* è il momento iniziale di ordine 1.
Il valore atteso è legato al valor medio della variabile, in quanto, per un gran numero di prove la media dei valori osservati di una variabile aleatoria tende (converge) alla sua speranza matematica.

SIMULAZIONE

Questa simulazione mostra come la media dei valori osservati di una variabile aleatoria distribuita secondo una determinata distribuzione tende, all'aumentare del numero di estrazioni generate, alla speranza matematica di tale distribuzione.
Per fare un esempio, consideriamo il lancio di una moneta.
Se la moneta non è truccata, dopo un discreto numero di lanci si dovrebbe ottenere lo stesso numero di teste e di croci: cosa che non si può dire in linea di principio dopo pochi lanci. Potrebbe infatti capitare che su tre lanci si presenti sempre una delle due facce.
Ovviamente da questo non possiamo dedurre che il risultato del lancio della moneta sarà sempre testa o sempre croce. Possiamo tuttavia sospettare della "bontà" della moneta se su un numero sufficientemente grande di lanci osserveremo una disparità significativa tra il numero di teste e quello di croci.

La funzione densità di probabilità è disegnata in **viola** come la speranza matematica. L'istogramma corrispondente alla realizzazioni della variabile aleatoria e disegnato in **rosso**, mentre la media campionaria è disegnata in **giallo**.

PROBABILITA`
- Definizioni -

Esistono diverse definizioni del concetto di *probabilità*: di fatto si possono elencare almeno quattro definizioni di probabilità, corrispondenti ad altrettanti approcci diversi che evidenziano ciascuno un aspetto di questo concetto.

Le quattro definizioni che andremo ad analizzare sono:

- Probabilità *Classica* o *Oggettiva*
- Probabilità *Empirica*
- Probabilità *Matematica*
- Probabilità *Soggettiva*

• LA PROBABILITA` CLASSICA

- Il calcolo delle probabilità è sorto con lo studio dei giochi d'azzardo all'incirca nel diciassettesimo secolo; proprio in questo periodo venne affermandosi la cosiddetta **definizione classica di probabilità**. Così come venne poi ripresa da C.F. Pierce nel 1910, la si può enunciare nel modo seguente:
- *la probabilità, P(A), di un evento A è il rapporto tra il numero N di casi "favorevoli" (cioè il manifestarsi di A) e il numero totale M di risultati ugualmente possibili e mutuamente escludentesi:*

$$P(A) = \frac{N}{M}$$

- Questa probabilità è talvolta detta *probabilità oggettiva* o anche *probabilità a priori*: il motivo dell'appellativo "a priori" deriva dal fatto che è possibile stimare la probabilità di un evento a partire dalla simmetria del problema.
  Ad esempio nel caso di un dado regolare si sa che la probabilità di avere un numero qualsiasi dei sei presenti sulle facce è 1/6, infatti, nel caso dell'uscita di un 3 si ha:

$$P(3) = \left(\frac{un\ 3}{una\ faccia\ qualsiasi\ del\ dado} \right) = \frac{1}{6}$$

- Analogamente nel caso di una puntata sul colore rosso alla roulette: la probabilità di vincere è pari a 18/37, circa il 49% (infatti i numeri rossi sono 18 su un totale di 37, essendoci oltre ai 18 numeri neri, anche lo zero che è verde).
- L'estensione al caso di eventi con risultati continui si attua attraverso una rappresentazione geometrica in cui la probabilità di un evento casuale è data dal rapporto tra l'area favorevole all'evento e l'area totale degli eventi possibili. Consideriamo un esempio.
- Supponiamo che un bambino lanci dei sassi contro una parete forata senza prendere la mira. Siano i fori sulla parete distribuiti a caso e per semplicità assumiamo che le dimensioni dei sassi siano molto piccole rispetto a quelle dei fori. Ci si chiede qual è la probabilità *p* che un sasso passi dall'altra parte.
  Se *A* è l'area della parete e *a* l'area di ciascuno dei *k* fori, la probabilità che un sasso passi è data dall'area "favorevole" divisa l'area totale

$$p = \frac{k\,a}{A}$$

- La quantità 1/A può essere considerata come una densità di probabilità. Essa è infatti la probabilità che ha un sasso di colpire una particolare superficie *unitaria* del muro: moltiplicando questa densità per l'area favorevole si ottiene direttamente la probabilità.
- Si noti bene che questa definizione oggettiva di probabilità diventa di difficile applicazione nelle numerose situazioni in cui la densità di probabilità non può più essere considerata uniforme, ovvero quando vengono meno le condizioni di simmetria. Ad esempio nel caso del bambino che lancia i sassi contro il muro può verificarsi che le dimensioni dei fori varino dal centro verso i bordi del muro e il bambino cerchi di mirare al centro.

LA PROBABILITA` EMPIRICA

La definizione di probabilità empirica è dovuta largamente a R. Von Mises ed è la definizione *sperimentale* di probabilità come **limite della frequenza** misurabile in una serie di esperimenti.

Essa ricalca lo schema della definizione classica, introducendo però un'importante variazione: sostituisce al rapporto *numero casi favorevoli / numero di casi possibili* il rapporto *numero di esperimenti effettuati con esito favorevole / numero complessivo di esperimenti effettuati*.
Il vantaggio di questa modifica è che questa definizione si applica senza difficoltà anche ai casi in cui la densità di probabilità non è uniforme, ovvero - per quanto riguarda esperimenti con risultati discreti - non è necessario specificare che i risultati debbano essere ugualmente possibili e mutuamente escludentesi.

Vediamo allora come viene definita la probabilità empirica:

> *la probabilità di un evento è il limite cui tende la frequenza relativa di successo all'aumentare del numero di prove.*

In pratica, se abbiamo un esperimento ripetuto *m* volte ed un certo risultato *A* che accade *n* volte, la probabilità di *A* è data dal limite della **frequenza** (*n/m*) quando m tende all'infinito

$$P(A) = \lim_{m \to \infty} \frac{n}{m}$$

Gli *m* tentativi possono essere effettuati sia ripetendo in sequenza *m* volte lo stesso esperimento sia misurando simultaneamente *m* esperimenti identici. L'insieme degli *m* tentativi prende il nome di **gruppo.**

Esistono alcune delicate questioni riguardo l'esistenza e il significato del limite presente nella definizione di probabilità empirica.

- La probabilità così definita non è una proprietà solo dell'esperimento, ma è anche funzione del particolare gruppo su cui viene calcolata. Ad esempio la probabilità di sopravvivenza ad una certa età, calcolata su diversi campioni di popolazione a cui una stessa persona appartiene (maschi, femmine, fumatori, non fumatori, deltaplanisti, ecc.), risulta diversa.
- La probabilità empirica si può rigorosamente applicare soltanto agli esperimenti ripetibili, per i quali il limite per *m* tende all'infinito ha significato. Si deduce quindi che molte situazioni della vita quotidiana, non sono soggette all'uso di questa definizione di probabilità: rappresentano esempi di questo tipo il risultato di una partita di calcio, quello di una roulette russa o il tempo atmosferico di domani.

Inoltre, per venire incontro ad una necessità di "operatività", quasi tutti sono concordi nel definire la probabilità come il valore della frequenza relativa di successo su un numero di prove non necessariamente tendente all'infinito, ma giudicato sufficientemente grande.

LA PROBABILITA` MATEMATICA

La definizione della probabilità matematica è dovuta a A.N. Kolmogorov.

> *Sia S un insieme di possibili risultati (A_i) di un esperimento*

$S = \{A_1, A_2, A_3, ...\}$ *se tali eventi sono [mutuamente escludentesi](#) allora per ognuno di essi esiste una probabilità P(A) rappresentata da un numero reale soddisfacente gli* ***assiomi di probabilità***

I) ***P(A)*** $\geq$ ***0*** *;*

II) se, come abbiamo ipotizzato, A_1 *e* A_2 *sono eventi mutuamente escludentesi, allora deve valere:*

$$P(A_1 \text{ oppure } A_2) = P(A_1) + P(A_2)$$

dove P (A_1 *oppure* A_2*) è la probabilità di avere il risultato* A_1 *o il risultato* A_2

III) $\sum_i P(A_i)$ ***= 1*** *, dove la sommatoria è estesa su tutti gli eventi mutuamente escludentesi.*

Le conseguenze dei tre assiomi sono:

- ***P (*** $\overline{A_i}$ ***) = P (non*** A_i ***) = 1 - P (*** A_i ***)***, cioè la probabilità di non ottenere l'evento A_i è uguale ad uno meno la probabilità di ottenerlo.
- ***P (*** A_i ***)*** $\leq$ ***1***, cioè la probabilità è un numero reale appartenente all'intervallo [0,1]: in pratica

  0 $\leq$ ***P (*** A_i ***)*** $\leq$ ***1***

  dove 1 rappresenta la certezza di ottenere l'evento A_i e 0 quella di non ottenerlo.

- La definizione di probabilità matematica può essere [estesa senza problemi alle variabili continue](#).
- Gli assiomi illustrati e le loro conseguenze sono sufficienti a fare qualsiasi conto, anche complicato, con le probabilità e ad ottenere un risultato numerico corretto. D'altra parte essi non dicono niente su cosa sia la probabilità e quale sia la sua natura: per questo si vedano le definizioni di [probabilità classica](#), [probabilità empirica](#) e [probabilità soggettiva](#).
- Tutte le definizioni di probabilità soddisfano le condizioni espresse da questi assiomi.

• LA PROBABILITA` SOGGETTIVA

- La concezione soggettivista della probabilità porta ad una definizione che si potrebbe formularsi nel modo seguente:
- *la probabilità di un evento A è la misura del grado di fiducia che un individuo coerente attribuisce, secondo le sue informazioni e opinioni, all'avverarsi di A.*
- Il campo di applicabilità di questa definizione è molto vasto; occorre aggiungere che la "coerenza" significa la corretta applicazione delle norme di calcolo e che la misura del grado di fiducia, cioè la valutazione di probabilità, viene fatta sull'intervallo (0,1) attribuendo la probabilità 0 all'evento impossibile e 1 a quello certo.

Circa il fatto che la valutazione sia qui un atto soggettivo, molto si potrebbe dire; interessa osservare che soggettivo non significa arbitrario. Di fatto appena vi siano premesse sufficienti, le valutazioni individuali concordano per quanti siano interessati al medesimo problema.

- Vediamo un esempio:
  Alessio è disposto a scommettere 1 contro 20 sul fatto che nel pomeriggio arrivi finalmente l'idraulico a riparare il rubinetto che perde da una settimana: attribuisce cioè a tale evento una probabilità pari ad 1/21 (meno del 5%). E` come se ci trovassimo ad effettuare un sorteggio da un'urna con 1 pallina rossa (evento positivo = arrivo dell'idraulico) e 20 palline nere (eventi negativi = assenza dell'idraulico).
- Ricapitolando: Alessio sta implicitamente dando una valutazione di probabilità, e tale valutazione attribuisce una probabilità 1/21 al realizzarsi dell'evento positivo = arrivo dell'idraulico. La frazione che esprime la probabilità ha numeratore uguale ad 1 che corrisponde a quanto Alessio è disposto a puntare e denominatore pari a 21 corrisponde alla puntata di Alessio sommata alla puntata di un ipotetico sfidante.
- Per quanto il dominio della probabilità soggettiva appaia incerto e arbitrario, vale la pena di osservare che proprio questa è la definizione di probabilità a cui più spesso ricorriamo nelle nostre considerazioni quotidiane ("domani pioverà, "questa volta passerò l'esame", ecc.).
- Esiste anche una definizione proposta da Bayes partendo dal teorema delle probabilità composte e dagli assiomi della probabilità.

• PROBABILITA`- Teoremi -

Per quanto riguarda il calcolo delle probabilità esistono alcuni teoremi fondamentali: elenchiamo quelli che approfondiremo.

- Teorema di addizione delle probabilità
- Teorema di moltiplicazione delle probabilità
- Teorema delle probabilità composte
- Formula della probabilità totale
- Teorema delle ipotesi (formula di Bayes)

Nell'illustrazione dei teoremi adotteremo le seguenti convenzioni:

1. $P(A \cap B) = P(A \times B) = P(A \text{ e } B)$ indica la probabilità che i due eventi A e B accadano contemporaneamente
2. $P(A \cup B) = P(A+B) = P(A \text{ oppure } B)$ equivale alla probabilità che si verifichi almeno uno dei due eventi o che si verifichino entrambe.

TEOREMA DELL'ADDIZIONE DELLE PROBABILITÀ

Attraverso questo teorema calcoliamo la probabilità di più eventi, sia che essi accadano insieme oppure in alternativa l'uno all'altro. Consideriamo due eventi: se essi sono mutuamente escludentesi allora la probabilità che accada o uno o l'altro è data dal secondo assioma della probabilità, mentre se l'accadere di uno non preclude la possibilità che si presenti anche il secondo allora bisogna ragionare in un altro modo.

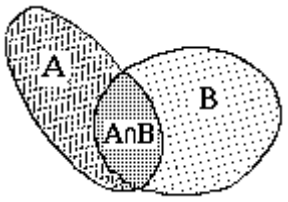
Nel caso che due eventi non siano mutuamente escludentesi, il teorema di addizione delle probabilità afferma che

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

che equivale a

$P(A+B) = P(A) + P(B) - P(A \times B)$

La dimostrazione del teorema si ricava osservando la figura accanto: se i due insiemi rappresentano rispettivamente la probabilità di successo dell'evento A e dell'evento B allora la probabilità che si verifichi almeno uno dei due è uguale alla somma delle aree dei due insiemi.
Però nel momento in cui si considera l'unione dei due insiemi bisogna togliere la quantità relativa alla loro intersezione in quanto viene considerata due volte: una volta per ciascun insieme.



Il teorema dell'addizione delle probabilità, come nel caso del prodotto, è facilmente estendibile a più di due eventi, sempre operando le stesse considerazioni che abbiamo fatto per i due eventi A e B.

TEOREMA DEL PRODOTTO DELLE PROBABILITA`

Per prodotto delle probabilità, in generale di due eventi si intende il verificarsi di entrambe **simultaneamente**. Ad esempio, se l'evento A consiste nell'estrazione di un due da un mazzo di carte e l'evento B nell'estrazione di una carta di picche, allora l'evento $C=A \times B$ consiste nell'apparizione di un due di picche.

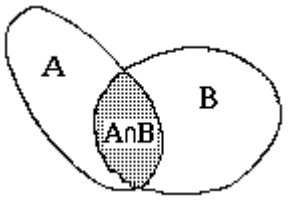
Il teorema si può quindi scrivere nella seguente forma:

$P(A \cap B) = P(A) + P(B) - P(A \cup B)$

che equivale a

$P(A \times B) = P(A) + P(B) - P(A+B)$

La dimostrazione del teorema è facilmente intuibile dall'osservazione della figura a fianco dove la quantità P(A∩B) corrisponde alla zona ombreggiata della figura. Tale zona, nel caso della somma dei due insiemi viene ricoperta due volte, quindi, togliendo la grandezza P(A∪B) si eliminano le parti non comuni dei due insiemi e in più si sottrae anche una volta la superficie P(A∩B) proprio come volevamo.



Il teorema del prodotto delle probabilità, come nel caso dell'addizione è facilmente estendibile al caso di più variabili applicando le stesse considerazioni fatte nel caso di due variabili.

Esiste anche un'altra formulazione di questo teorema che va sotto il nome di teorema delle probabilità composte.

TEOREMA DELLE PROBABILITÀ COMPOSTE

Un altro modo per esprimere il teorema di moltiplicazione delle probabilità è rappresentato dal teorema delle probabilità composte. Tale teorema può essere enunciato nel modo seguente:

> *la probabilità del prodotto di due eventi è uguale al prodotto della probabilità di uno degli eventi per la probabilità condizionata dell'altro calcolata a condizione che il primo abbia luogo.*
> P(A∩B) = P(A) P(B|A) = P(B) P(A|B)

Si noti che se gli eventi A e B sono mutuamente escludentesi, la probabilità condizionata si annulla per definizione, mentre se gli eventi sono indipendenti si ha che P(B|A) = P(B) e analogamente P(A|B) = P(A): da quest'ultimo caso discende un **corollario** molto importante:

> *la probabilità del prodotto di due eventi indipendenti è uguale al prodotto delle probabilità di questi eventi.*
> P(A∩B) = P(A) P(B)
>
> *o in generale, per N eventi*

$$P\left(\prod_{i=1}^{N} A_i\right) = \prod_{i=1}^{N} P(A_i)$$

Esempio

Un'urna contiene 2 palline bianche e 3 palline rosse. Si estraggono due palline dall'urna: trovare la probabilità che le due palline estratte siano entrambe bianche.

Denotiamo con A l'evento - apparizione di due palline bianche -. L'evento A è il prodotto di due eventi elementari

$A = A_1 A_2$

dove A_1 rappresenta l'apparizione della pallina bianca alla prima estrazione e A_2 è la probabilità di ottenere la pallina bianca alla seconda estrazione.
In virtù del teorema delle probabilità composte

$$P(A) = P(A_1)P(A_2|A_1) = (2/5) \times (1/4) = 0.1$$

Consideriamo la stessa situazione in cui dopo la prima estrazione, la pallina estratta viene rimessa nell'urna (si parla allora di *estrazione con reintroduzione*): in questo caso gli eventi A_1 e A_2 sono indipendenti e per il corollario si ha:

$$P(A) = P(A_1)P(A_2) = (2/5) \times (2/5) = 0.16$$

FORMULA DELLA PROBABILITÀ TOTALE

La **formula della probabilità totale** è un corollario di due teoremi fondamentali quali il teorema di addizione e il teorema di moltiplicazione delle probabilità.

Si chiede di determinare la probabilità di un certo evento *X* che può verificarsi insieme ad uno degli eventi

$$\{A_1, A_2, A_3, ...\}$$

che formano un gruppo completo di eventi mutuamente escludentesi. Chiameremo questi eventi *ipotesi*.
Dimostriamo allora che in questo caso si ha:

$$P(X) = \sum_{i=1}^{N} P(A_i)P(X|A_i) \qquad (1)$$

vale a dire che la probabilità dell'evento *X* si calcola come la somma dei prodotti delle probabilità di ciascuna delle ipotesi per la probabilità condizionata dell'evento con tali ipotesi.

Formando le ipotesi A_i un gruppo completo, l'evento *X* non si verifica che in combinazione con una di queste ipotesi

$$X = A_1X + A_2X + ... + A_NX$$

Essendo le ipotesi incompatibili, lo sono ugualmente le singole combinazioni A_i*X*; applicando loro il teorema di addizione, si ottiene

$$P(X) = P(A_1X) + P(A_2X) + ... + P(A_NX) = \sum_{i=1}^{N} P(A_iX)$$

Applicando all'evento A_i X il teorema di moltiplicazione si ottiene la **(1)**.

TEOREMA DELLE IPOTESI
- FORMULA DI BAYES -

Il teorema delle ipotesi è un corollario del teorema di moltiplicazione e della formula della probabilità totale.

Consideriamo il seguente problema: sia un gruppo di eventi (che in seguito chiameremo ipotesi) mutuamente escludentesi $\{A_1, A_2, A_3, ...\}$.

Siano inoltre note le probabilità delle singole ipotesi e valgano rispettivamente P(A_1), P(A_2), ...,P(A_N).

Si esegue un esperimento durante il quale si verifica un certo evento X. Ci si domanda: in che modo si devono cambiare le probabilità delle ipotesi in seguito al verificarsi di questo evento?

In pratica si vuol trovare la probabilità condizionata P(A_i | X) per ciascuna ipotesi.

In accordo con il teorema di moltiplicazione abbiamo

$$P(X A_i) = P(X)P(A_i|X) = P(A_i)P(X|A_i)$$ o, eliminando il primo membro

$$P(X)P(A_i|X) = P(A_i)P(X|A_i)$$

da cui possiamo ricavare

$$P(A_i|X) = \frac{P(A_i)P(X|A_i)}{P(X)}$$

Ma dalla formula della probabilità totale sappiamo che

$$P(X) = \sum_{i=1}^{N} P(A_i)P(X|A_i) \qquad (1)$$

per cui sostituendo nella formula precedente otteniamo

$$P(A_i|X) = \frac{P(A_i)P(X|A_i)}{\sum_{i=1}^{N} P(A_i)P(X|A_i)}$$

Quest'ultima è detta **formula di Bayes** o **teorema delle ipotesi**.

TEOREMA DELLE IPOTESI
- Esempi -

- **Esempio 1**. Due tiratori sparano un colpo ciascuno sul medesimo bersaglio.
  La probabilità che il primo tiratore colpisca il bersaglio è 0.8 (80%), mentre quella del secondo tiratore è 0.4 (40%). Il bersaglio viene colpito una sola volta.
  Trovare la probabilità che il bersaglio sia stato colpito dal primo tiratore.

  Soluzione. Prima di sparare i colpi sono possibili le seguenti *ipotesi*:

  A_1 - Entrambi i tiratori mancano il bersaglio
  A_2 - Entrambi i tiratori centrano il bersaglio
  A_3 - Il primo tiratore centra il bersaglio, mentre il secondo lo manca
  A_4 - Il primo tiratore manca il bersaglio, mentre il secondo lo centra

  Le probabilità di queste ipotesi valgono:

  $P(A_1) = 0.2 \times 0.6 = 0.12$
  $P(A_2) = 0.8 \times 0.4 = 0.32$
  $P(A_3) = 0.8 \times 0.6 = 0.48$
  $P(A_4) = 0.2 \times 0.4 = 0.08$

  Le probabilità condizionate del verificarsi dell'evento *X* (bersaglio colpito una sola volta) per queste ipotesi sono uguali a:

  $P(X \mid A_1) = 0$
  $P(X \mid A_2) = 0$
  $P(X \mid A_3) = 1$
  $P(X \mid A_4) = 1$

  Dopo la prova le ipotesi A_1 e A_2 diventano impossibili, e le probabilità delle ipotesi A_3 e A_4 per la formula di Bayes sono:

$$P(A_3|X) = \frac{0.48 \cdot 1}{0.48 \cdot 1 + 0.08 \cdot 1} = \frac{6}{7}$$

$$P(A_4|X) = \frac{0.08 \cdot 1}{0.48 \cdot 1 + 0.08 \cdot 1} = \frac{1}{7}$$

  Di conseguenza la probabilità che il bersaglio sia stato colpito dal primo tiratore e pari a 6/7.

- **Esempio 2**. Due stazioni di osservazione forniscono dei dati su un certo oggetto che può trovarsi in due stati *M* e *N*, passando in modo aleatorio dall'uno all'altro.
  Lunghe osservazioni hanno permesso di stabilire che durante circa il 30% del tempo l'oggetto si trova nello stato *M* e per il restante 70% nello stato *N*.
  La stazione numero 1 fornisce dei dati erronei per circa il 2% dei casi totali, mentre la stazione numero 2 per l'8%. Ad un certo punto la stazione n.1 comunica che l'oggetto si trova nello stato *M* e la stazione n.2 contemporaneamente lo avvista nello stato *N*.
  Ci si domanda quale delle due comunicazioni deve essere supposta corretta.

  Soluzione. E` naturale supporre vera la comunicazione la cui probabilità di errore è minore: vediamo cosa accade applicando la formula di Bayes. A tale scopo impostiamo le ipotesi riguardanti lo stato dell'oggetto:

  A_1 - L'oggetto si trova nello stato *M*
  A_2 - L'oggetto si trova nello stato *N*

  In questo caso l'evento osservato *X* corrisponde al fatto seguente: la stazione n.1 ha trasmesso che l'oggetto si trova nello stato *M* e la stazione n.2 che esso si trova nello stato *N*.
  Le probabilità delle ipotesi prima dell'avvistamento sono:

  $P(A_1) = 0.3$
  $P(A_2) = 0.7$

  Calcoliamo le probabilità condizionate dell'evento osservato *X* per queste ipotesi.
  Affinché l'evento *X* abbia luogo per l'ipotesi A_1 è necessario che la comunicazione trasmessa dalla prima stazione sia vera e quella proveniente dalla seconda sia errata, cioè:

  $P(X \mid A_1) = 0.98 \times 0.08 = 0.0784$

  Analogamente

  $P(X \mid A_2) = 0.92 \times 0.02 = 0.0184$

  Applicando la formula di Bayes, troviamo la probabilità che lo stato reale dell'oggetto sia *M*:

$$P(A_1|X) = \frac{0.3 \cdot 0.0784}{0.3 \cdot 0.0784 + 0.7 \cdot 0.0184} \approx 0.645$$

  Cioè che la comunicazione della prima stazione fosse quella corretta.

DISTRIBUZIONI DI PROBABILITA`

Dagli assiomi della probabilità introdotti nella definizione matematica di probabilità data da A.N.Kolmogorov, sappiamo che, dato un insieme di possibili valori mutuamente escludentesi di una variabile aleatoria, per il terzo assioma si ha:

$$\sum_i P(A_i) = 1$$

Cioè la somma delle probabilità di tutti i valori possibili di una variabile aleatoria è uguale all'unità. Questa probabilità totale si *distribuisce* in un certo modo tra i diversi valori della variabile.
Al fine di descrivere una variabile aleatoria dal punto di vista probabilistico specifichiamo questa distribuzione, cioè indichiamo esattamente la probabilità di ciascuno dei valori possibili. Si stabilisce così la **legge distribuzione** della variabile aleatoria.
Allora possiamo concludere che:

> *si definisce **legge di distribuzione** di una variabile aleatoria ogni relazione che stabilisce una corrispondenza tra i valori possibili di tale variabile e la loro probabilità.*

Il fatto che una variabile aleatoria si distribuisca secondo una data distribuzione ci permette di trarre alcune conclusioni importanti tra cui la possibilità di definire quello che viene comunemente chiamato **livello di confidenza**: il livello di confidenza non è altro che la probabilità che l'affermazione a cui esso si riferisce sia vera.

Vediamo alcuni casi di distribuzioni per due differenti tipi di variabili aleatorie:

- distribuzioni per variabili discrete
- distribuzioni per variabili continue

DISTRIBUZIONI DISCRETE

Esistono svariati tipi di distribuzioni per le variabili aleatorie discrete. Quelle che andremo ad analizzare sono:

- distribuzione uniforme
- distribuzione binomiale
- distribuzione di Poisson
- distribuzione ipergeometrica

Va tenuto presente che alcune di queste saranno approssimabili con altre distribuzioni continue nel momento il cui il numero di prove effettuate diventa abbastanza elevato.

DISTRIBUZIONI DISCRETE
- Distribuzione uniforme -

Una variabile aleatoria discreta tipica è il lancio di un dado regolare: i sei eventi possibili sono i sei possibili punteggi ottenibili con il lancio, a ciascuno dei quali corrisponde il valore di probabilità

$$P_i = \frac{1}{6}$$

in accordo con il terzo assioma della probabilità:

$$\sum_{i=1}^{6} P_i = 1$$

In generale una distribuzione uniforme per una variabile aleatoria discreta, capace di assumere *N* valori, ha la forma seguente:

$$P_i = \frac{1}{N}$$

La media di tale distribuzione si trova sfruttando il momento iniziale di ordine 1, per cui:

$$\overline{m} = \sum_{i=1}^{N} i P_i = \sum_{i=1}^{N} \frac{i}{N} = \frac{1}{N} \sum_{i=1}^{N} i = \frac{1}{N} \frac{N(N+1)}{2} = \frac{N+1}{2}$$

Per quanto riguarda la deviazione standard, utilizzando la definizione di varianza:

$$D = \alpha_2 - \overline{m}^2 = \sum_{i=1}^{N} \frac{i^2}{N} - \left(\frac{N+1}{2}\right)^2 = \frac{1}{N} \sum_{i=1}^{N} i^2 - \left(\frac{N+1}{2}\right)^2 =$$

$$= \frac{1}{N} \frac{N(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} = \frac{N^2-1}{12}$$

E la deviazione standard risulta essere:

$$\sigma_m = \sqrt{\frac{N^2-1}{12}}$$

DISTRIBUZIONI DISCRETE
- Distribuzione binomiale -

Per introdurre la distribuzione binomiale ricorriamo ad un esempio.
Supponiamo che tre persone (**F**rancesca, **L**uigi e **T**iziano) escono ciascuno dalla loro casa per andare a prendere il medesimo autobus e che ciascuno di essi abbia probabilità pari a *p* di riuscire

ad arrivare in tempo alla fermata (e ovviamente probabilità *1-p* di perdere l'autobus): ci si chiede quale sia la probabilità che due dei tre personaggi in questione riesca nell'intento.

Cominciamo col notare che si richiede la probabilità che due prendano l'autobus, senza specificare quali: in questo modo l'evento *due persone prendono l'autobus* si può verificare in tre modi diversi ossia

1) Francesca e Luigi lo prendono, ma Tiziano no
2) Francesca e Tiziano lo prendono, ma Luigi no
3) Il terzo caso è facilmente intuibile...

Dunque l'evento *almeno in due prendono l'autobus*, che denoteremo B_2, sarà rappresentabile come

$$B_2 = FL\overline{T} + F\overline{L}T + \overline{F}LT$$

dove **F**, **L** e **T** sono rispettivamente Francesca, Luigi e Tiziano che prendono l'autobus, mentre $\overline{F}$, $\overline{L}$ e $\overline{T}$ corrispondono ognuno al rispettivo personaggio deluso per aver perso l'autobus.
Potendo inoltre considerare i tre personaggi (eventi) indipendenti, per i teoremi della somma e del prodotto delle probabilità, la probabilità dell'evento B_2 sarà:

$$P(B_2) = pp(1-p) + p(1-p)p + (1-p)pp$$

Ponendo *1 - p = q* otteniamo in definitiva

$P(B_2) = 3\,p^2q$

Vediamo la generalizzazione di questo esempio.
Si considerino *N* prove indipendenti in cui l'evento *A* può verificarsi o meno: sia *p* la probabilità (**costante** per ogni prova) che l'evento *A* si presenti e di conseguenza *1-p=q* la probabilità che esso non si verifichi. Cerchiamo la probabilità $P_{m,N}$ che l'evento *A* si verifichi *m* volte in *N* prove.

Consideriamo a questo proposito l'evento B_m corrispondente al verificarsi di *A* esattamente *m* volte in *N* prove.
Come nell'esempio precedente questo evento può realizzarsi in più modi diversi: decomponiamo allora l'evento B_m in una somma di prodotti di eventi, consistenti nel presentarsi o meno di *A* in una singola prova.
Se denotiamo con A_i il presentarsi dell'evento *A* nell'i-esima prova e con $\overline{A_i}$ il non presentarsi di *A* nell'i-esima prova, abbiamo che ogni variante di apparizione dell'evento B_m si compone di *m* apparizioni dell'evento *A* e di *n-m* eventi $\overline{A}$ con indici distinti

$$B_m = A_1A_2\ldots A_m\overline{A_{m+1}}\ldots\overline{A_N} + A_1\overline{A_2}A_3\ldots\overline{A_{N-1}}A_N + \ldots + \overline{A_1A_2}\ldots\overline{A_{N-m}}A_{N-m+1}\ldots A_N$$

Il numero di combinazioni possibili è uguale a C_N^m, cioè al numero di modi diversi in cui si possono scegliere le *m* prove, tra le *N* totali, in cui abbia luogo l'evento *A*.
Per il teorema di moltiplicazione delle probabilità nel caso di eventi indipendenti, la probabilità di ogni combinazione è $p^m q^{N-m}$.
Essendo le varie combinazioni mutuamente escludentesi, per il teorema dell'addizione, la probabilità dell'evento B_m è pari a :

$$P_{m,N} = \underbrace{p^m q^{N-m} + \ldots + p^m q^{N-m}}_{C_N^m \text{ volte}} = C_N^m p^m q^{N-m}$$

Il coefficiente C_N^m è detto *coefficiente binomiale* e viene spesso indicato come $\binom{N}{m}$: in particolare il suo valore è

$$C_N^m = \frac{N!}{m!(N-m)!}$$

DISTRIBUZIONI DISCRETE
- Distribuzione binomiale -

Note

Il coefficiente binomiale che compare nell'omonima distribuzione deriva dall'espansione del *binomio di Newton* $(p+q)^N$ in $N+1$ termini: tale sviluppo si può scrivere come

$$(p+q)^N = p^N + Np^{N-1}q + \ldots + Npq^{N-1} + q^N = \sum_{m=0}^{N} \binom{N}{m} p^m q^{N-m}$$

e vale per due numeri qualunque *p* e *q* e ogni intero positivo *n*.

Proprietà della distribuzione binomiale

La distribuzione binomiale dà la probabilità di ottenere *m* successi in *n* prove, quando *p* è la probabilità di successo in una singola. Il valor medio di tale distribuzione, corrispondente al numero medio di successi, si ricava attraverso il momento iniziale di ordine 1:

$$\overline{m} = \sum_{m=0}^{N} m \binom{N}{m} p^m q^{N-m} = \sum_{m=1}^{N} \frac{N! m}{(N-m)! m!} p^m q^{N-m} =$$

$$= Np \sum_{m=1}^{N} \frac{(N-1)!}{[(N-1)-(m-1)]!\,(m-1)!} p^{m-1} q^{N-m} =$$

Ponendo m - 1 = r otteniamo

$$= Np \sum_{r=0}^{N-1} \frac{(N-1)!}{(N-1-r)!\,r!} p^r q^{N-1-r} = Np$$

Infatti l'ultima sommatoria è equivalente all'espansione del binomio di Newton

$$\sum_{m=0}^{N} \binom{N}{m} p^m q^{N-m} = (p+q)^N = 1^N = 1$$

considerato nel nostro caso in cui la somma p+q è uguale ad uno.
Così come abbiamo ricavato il valor medio di successi in *N* prove, attraverso il momento centrale di ordine 2 si può ricavare la deviazione standard.
Si ottiene così come valore della deviazione standard:

$$\sigma_m = \sqrt{Np(1-p)}$$

In generale la distribuzione binomiale non è simmetrica salvo il caso in cui, tipico ad esempio del lancio della moneta, p sia uguale ad 1/2.

 Approfondimento

---

Esempi

- **Esempio 1**. Cinque stazioni radio sono in comunicazione tra loro. Ogni tanto la comunicazione si interrompe a causa delle condizioni meteorologiche.
  Essendo le stazioni molto lontane tra di loro consideriamo ragionevolmente che gli arresti delle comunicazioni siano indipendenti tra loro e inoltre che si producano con una probabilità *p=0.2*.
  Trovare la probabilità che a un dato istante la comunicazione venga mantenuta da *almeno* due delle cinque stazioni radio.

  Soluzione. In questo caso bisogna prestare attenzione ad un particolare: in questo esempio ci viene richiesta la probabilità che **almeno** due stazioni siano funzionanti. Per fare questo applichiamo il teorema della somma alle probabilità dei seguenti eventi:

  non vi è alcuna interruzione ($P_{0,5}$)
  interruzione dovuta ad 1 stazione ($P_{1,5}$)
  interruzione dovuta ad 2 stazioni ($P_{2,5}$)
  interruzione dovuta ad 3 stazioni ($P_{3,5}$)

  La somma delle probabilità di questi quattro eventi ci dà la probabilità che cerchiamo (***P***) ossia che almeno due stazioni siano funzionanti.

$$P = P_{0,5} + P_{1,5} + P_{2,5} + P_{3,5} = C_5^0 p^5 + C_5^1 p^4 q + C_5^2 p^3 q^2 + C_5^3 p^2 q^3 =$$

= 0.0003+0.0064+0.0512+0.2048=0.2627

Cioè la probabilità che almeno due stazioni mantengano la comunicazione ad un dato istante è circa il 26%.

- **Esempio 2**.Un'azienda decide di verificare il funzionamento di una macchina di sua proprietà che produce un determinato articolo. Per fare questo gli incaricati esaminano gruppetti di oggetti sfornati dalla macchina costituiti da 5 articoli. Il risultato del prelievo dei campioni è riassunto nella seguente tabella:

Articoli difettosi (W)	0	1	2	3	4	5
N campioni con W articoli difettosi	58	32	7	2	1	0

Mostrare che la distribuzione degli articoli fallati riportata nella tabella è assimilabile ad una distribuzione di tipo binomiale e determinare la percentuale di articoli prodotti dalla macchina e affetti da qualche difetto.

Soluzione. Dalla tabella riportata calcoliamo quanti articoli difettosi sono stati raccolti in 100 campioni e, di seguito, dividendo per il numero totale dei campioni prelevati, la media della distribuzione empirica riportata in tabella:

(0+32+14+6+4+0) /100 = 0.56

Se ora supponiamo che tale distribuzione sia modellizzabile attraverso una distribuzione binomiale, il valore medio di articoli difettosi per ogni campione che ci aspetterebbe è **Np**: qui **N** è uguale al numero di articoli per ogni campione cioè 5, per cui il calcolo di **p** (probabilità di trovare un articolo difettoso in un campione di cinque oggetti) è immediato e si ricava:

Np=0.56 porta, nel caso di **N**=5 a **p=0.112** che possiamo ben approssimare con **p=1/9**

Ovviamente in questo caso **q=8/9** e con questi valori di **p** e **q** andiamo a calcolare lo sviluppo del binomio di Newton $(p+q)^N$ che ci da la distribuzione binomiale che noi abbiamo ipotizzato:

$$\left(\frac{1}{9} + \frac{8}{9}\right)^5 = \frac{1}{9^5}\left(8^5 + 5 \cdot 8^4 + 10 \cdot 8^3 + 10 \cdot 8^2 + 5 \cdot 8 + 1\right) =$$

= 0.5549+0.3468+0.0867+0.0108+0.0007+0.0000

in questo modo, la nostra distribuzione ipotetica ci dovrebbe dare, su 100 campioni di cinque oggetti ciascuno, 55 campioni privi di articoli difettosi, 35 con un articolo difettoso, 9 con due articoli difettosi, 1 con tre oggetti fallati e nessun campione con quattro o cinque articoli difettosi.

I dati così ottenuti si conformano alla distribuzione sperimentale trovata dagli addetti ai controlli: possiamo quindi riassumere dicendo che la distribuzione è approssimativamente

binomiale e la percentuale di articoli difettosi prodotti dalla macchina è stimabile con il valore di **p** che abbiamo ottenuto, cioè circa 11%.

DISTRIBUZIONI DISCRETE
- Distribuzione di Poisson -

La distribuzione di Poisson si può ricavare come caso particolare della distribuzione binomiale: quando cioè il numero di prove **N** diventa molto grande e contemporaneamente la probabilità di successo in una singola prova molto piccola, in modo tale che il loro prodotto sia finito (non diverga) e diverso da zero.
Una tipica situazione che gode delle caratteristiche appena elencate è lo studio di un processo di decadimento radioattivo. In questa circostanza, il numero di prove è costituito dal numero di nuclei che potenzialmente possono decadere (per una mole di materiale radioattivo il numero di nuclei è dell'ordine di 10^{23}), mentre la probabilità di "successo" (decadimento) per ogni nucleo è decisamente piccola.
Si suppone che la probabilità **p** di decadimento sia costante e inoltre, altra caratteristica tipica dei cosiddetti processi di Poisson, che la probabilità di successo in un intervallo di tempo $[t, t+\Delta t]$ sia proporzionale, in prima approssimazione, a Δt.

Vediamo ora la forma matematica di tale distribuzione. Una variabile aleatoria si distribuisce in modo poissoniano se la probabilità di ottenere *m* successi è data da:

$$P_m = \frac{a^m}{m!} e^{-a}$$

dove la grandezza *a* è detta *parametro* della legge di Poisson e rappresenta la *frequenza media* di accadimento dell'evento osservato.
Ad esempio, la probabilità di ottenere due successi è data da $e^{-a}a^2/2!$, quella di ottenerne tre è $e^{-a}a^3/3!$ e via dicendo.

Note

La distribuzione Poisson è in accordo terzo assioma della probabilità. Facendo la sommatoria su tutti gli eventi possibili otteniamo:

$$\sum_{m=0}^{\infty} P_m = \sum_{m=0}^{\infty} \frac{a^m}{m!} e^{-a} = e^{-a} \sum_{m=0}^{\infty} \frac{a^m}{m!}$$

Ma la serie in *m* è proprio la serie che definisce e^a per cui

$$\sum_{m=0}^{\infty} P_m = e^{-a} \cdot e^{a} = 1$$

Proprietà della distribuzione di Poisson

Vediamo qual'è il valor medio della distribuzione di Poisson avvalendoci del momento iniziale di ordine 1:

$$\overline{m} = \sum_{m=0}^{\infty} m \frac{a^m}{m!} e^{-a} = ae^{-a} \sum_{m=1}^{\infty} \frac{ma^{m-1}}{m!} = ae^{-a} \sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!}$$

Si noti che ora la sommatoria parte da 1 in quanto il termine corrispondente a m=0 è nullo. Se poniamo $m-1=k$ l'ultima serie è di nuovo la sommatoria che definisce e^a, per cui:

$$\overline{m} = ae^{-a} \sum_{k=0}^{\infty} \frac{a^k}{(k)!} = ae^{-a}e^a = a$$

Ricorrendo alla definizione del momento centrale di ordine 2, si può dimostrare che la deviazione standard è uguale alla radice quadrata del *parametro* caratteristico della distribuzione di Poisson

$$\sigma_m = \sqrt{a}$$

Esempi

- **Esempio 1**. Supponiamo di voler calcolare la probabilità che una qualsiasi persona adulta abbia bisogno del medico nell'arco di un anno. Utilizziamo a questo scopo alcuni dati rilevati a Philadelphia nel 1967 e riportati nella tabella seguente:

Visite mediche, Philadelphia 1967

N. visite	N. persone	Tot. visite
0	40	0
1	30	30
2	15	30
3	8	24
4	4	16
5	2	10
6	1	6
0-6	100	116

Queste cifre si adattano alla distribuzione di Poisson: il numero di persone che va dal medico è positivo e inoltre è *piccolo* rispetto al totale del campione (il massimo della distribuzione è proprio sullo 0!).

Dalla tabella risulta che $\overline{m}=a$=116/100=1.16: ne segue che la probabilità che una persona a caso non abbia mai bisogno del medico nel corso di un anno è: $P_0=e^{-1.16}$=0.313;. Per la prima conseguenza degli assiomi della probabilità, la probabilità che una persona a caso vada dal medico *almeno* una volta in un anno è $P_{\geq 1}=1-P_0=0.687$, etc.

- **Esempio 2**. Calcolare quanti conteggi di decadimenti radioattivi bisogna effettuare per ottenere una precisione del 5%.

  Soluzione. Se si contano *N* eventi rari, frutto di uno steso processo aleatorio con probabilità costante, l'errore statistico da associare a tale numero è $\sigma = \sqrt{N}$ in questo modo l'errore relativo o percentuale decresce, all'aumentare di *N*, come $\frac{1}{\sqrt{N}}$.
  Quindi per realizzare una misura precisa è necessario aumentare adeguatamente il numero di conteggi da effettuare: in particolare, per ottenere una precisione del 5% occorrerà che sia verificato

  $$\frac{\sqrt{N}}{N} = 0.05$$

  Risolvendo questa equazione otteniamo il numero di conteggi necessario per ottenere una precisione del 5%, ossia *N*=400.

- **Esempio 3**. Ad un centralino arrivano delle chiamate con una densità media *k* ogni ora. Supponendo che il numero di chiamate in un qualsiasi intervallo di tempo sia distribuito secondo la legge di Poisson, calcolare la probabilità con la quale arrivano esattamente 3 chiamate in 2 minuti.

  Soluzione. Il numero medio di chiamate in 2 minuti, dal valore medio sull'intervallo di un'ora, è uguale a:

  $$a = \frac{2k}{60} = \frac{k}{30}$$

  Applicando la formula della distribuzione di Poisson, otteniamo per *m*=3 chiamate:

  $$P_3 = \frac{\left(\frac{k}{30}\right)^3}{1 \cdot 2 \cdot 3} e^{-k/30}$$

DISTRIBUZIONI DISCRETE
- Distribuzione ipergeometrica -

La distribuzione ipergeometrica si applica ad insiemi contenenti *N* elementi divisi in due classi: in una abbiamo *M* oggetti che presentano una certa caratteristica e nell'altra si trovano gli altri *N*-*M* elementi con caratteristiche diverse da quella che contraddistingue gli oggetti della prima classe.
Se ci si chiede quale sia la probabilità di trovare *x* elementi appartenenti alla prima classe, effettuando un'estrazione, senza reintroduzione, di campione a caso di *n* elementi dagli *N* totali, bisogna utilizzare alla distribuzione ipergeometrica.

La forma matematica di tale distribuzione è la seguente:

$$P_x = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Si noti bene che a differenza del caso della distribuzione binomiale, ciascuna delle *n* estrazioni **non** sono indipendenti l'una dalle altre.

Note

Ricordiamo che con la scrittura $\binom{N}{m}$ indichiamo il *coefficiente binomiale* C_N^m

$$C_N^m = \frac{N!}{m!(N-m)!}$$

Una proprietà dei coefficienti binomiali che ci sarà utile per i calcoli successivi è la seguente:

$$\sum_{x=0}^{n} \binom{M}{x} \binom{N-M}{n-x} = \binom{N}{n}$$

Proprietà della distribuzione ipergeometrica

Calcoliamo il valor medio della distribuzione utilizzando la definizione del momento iniziale di ordine 1:

$$\overline{m} = \sum_{x=0}^{n} x \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \sum_{x=1}^{n} M \frac{\binom{M-1}{x-1} \binom{N-M}{n-x}}{\binom{N}{n}}$$

ponendo $y = x - 1$ otteniamo

$$\overline{m} = M \sum_{y=0}^{n-1} \frac{\binom{M-1}{y} \binom{N-M}{n-1-y}}{\binom{N}{n}}$$

Dalla proprietà dei coefficienti binomiali, otteniamo

$$\overline{m} = M \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = M \cdot \frac{n}{N}$$

Per quanto riguarda la deviazione standard, dopo lunghi e tediosi calcoli si arriva a

$$\sigma_m = \sqrt{\frac{M}{N} \cdot n \cdot \frac{(N-M)(N-n)}{N(N-1)}}$$

Approfondimenti

Esempi

- **Esempio 1**. Un'urna contiene N=10 palline: M=7 di queste sono di colore bianco e le rimanenti tre sono nere.
  Calcolare la probabilità che estraendo un campione a caso di n=5 palline contenente, x=3 di queste siano bianche.

  Soluzione. Applicando la formula della distribuzione ipergeometrica con i dati forniti nel testo otteniamo

$$P_3 = \frac{\binom{7}{3}\binom{10-7}{5-3}}{\binom{10}{5}} = \frac{\frac{7!}{3!4!} \cdot \frac{3!}{2!1!}}{\frac{10!}{5!5!}} = \frac{105}{252}$$

  Cioè la probabilità di estrarre un campione di 5 palline contenente 3 palline bianche è pari a 105/252 (=2/5) ossia poco meno del 42%.

- **Esempio 2**. Con gli stessi dati dell'esempio precedente si calcoli la probabilità di estrarre un campione contenente **almeno** due palline bianche.

  Soluzione . Per calcolare la probabilità che almeno 2 palline del campione siano bianche, utilizziamo il teorema della somma delle probabilità considerando un campione con almeno due palline bianche.
  In particolare i casi possibili sono: 2 palline bianche su 5, 3 su 5, 4 su 5 oppure tutte e 5 bianche. allora la probabilità richiesta sarà

$$P_{Tot} = P_2 + P_3 + P_4 + P_5 = \frac{\binom{7}{2}\binom{3}{3}}{\binom{10}{5}} + \frac{\binom{7}{3}\binom{3}{2}}{\binom{10}{5}} + \frac{\binom{7}{4}\binom{3}{1}}{\binom{10}{5}} + \frac{\binom{7}{5}\binom{3}{0}}{\binom{10}{5}}$$

  Eseguendo i calcoli otteniamo

$$P_{Tot} = \frac{21}{252} + \frac{3 \cdot 35}{252} + \frac{3 \cdot 35}{252} + \frac{21}{252} = \frac{252}{252} = 1$$

  Questo risultato non deve stupirci in quanto, essendo solo 3 le palline nere contenute nell'urna, al momento dell'estrazione di 5 palline per forza almeno 2 palline dovranno essere bianche!!!

  Se infatti ci chiedessimo quale sia la probabilità di estrarre un campione di 5 palline contenente solo una pallina bianca, ci troveremmo di fronte al fattoriale di un numero negativo che sappiamo essere privo di significato (l'esercizio è lasciato allo studente).

DISTRIBUZIONI CONTINUE
- Introduzione -

Nel momento in cui si parla di distribuzioni di probabilità per variabili aleatorie continue, bisogna prestare un attimo di attenzione ad un'importante questione. Attribuendo alla variabile aleatoria la potenza del continuo, accade che in un qualsiasi intervallo finito cadano un'infinità di valori della variabile stessa: di conseguenza non si può pensare di attribuire a ciascuno di essi un valore finito della probabilità.
Per questo motivo introduciamo alcuni concetti fondamentali che in seguito ci saranno molto utili per lo studio delle distribuzioni di variabili continue.

Funzione di distribuzione cumulativa

La *funzione di distribuzione cumulativa* per una variabile aleatoria continua *X* è definita come la probabilità che la variabile *X* assuma un qualsiasi valore minore di un valore *x*:

$$P(X < x) = F(X)$$

La funzione di distribuzione cumulativa è una caratteristica di una variabile aleatoria. Essa esiste per tutte le variabili aleatorie, siano esse discrete o continue.
Vediamone ora alcune *proprietà fondamentali*:

1) La funzione cumulativa *F(x)* è una funzione **non decrescente**, vale a dire che per $x_2 > x_1$ si ha $F(x_2) \geq F(x_1)$.
2) Quando l'argomento *x* della funzione tende a $-\infty$ la funzione di distribuzione tende a zero: **$F(-\infty) = 0$**
3) Quando invece l'argomento *x* tende a $+\infty$ la funzione di distribuzione tende a uno: **$F(+\infty) = 1$**

Senza dare una dimostrazione rigorosa di queste proprietà vediamo come esse siano di facile comprensione attraverso un esempio: esempio che, per facilitare la comprensione, viene presentato inizialmente per variabili discrete.
Supponiamo di avere una variabile aleatoria discreta che può assumere solo cinque valori: le probabilità di ottenere i singoli valori sono raccolte nella tabella seguente.

x_i	0	1	2	3	4
p_i	0.15	0.37	0.28	0.11	0.09

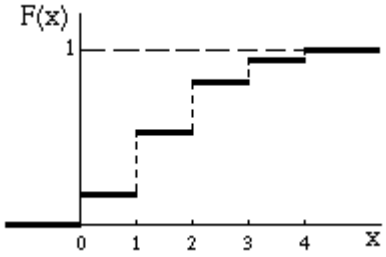
Andiamo ora a "costruire" la funzione di distribuzione cumulativa, tenendo presente che

$$F(X) = P(X < x) = \sum_{x_i < x} P(X = x_i)$$

dove la disuguaglianza $x_i < x$ sotto il segno di sommatoria significa che questa è estesa a tutti gli x_i inferiori ad *x*. Per cui abbiamo

$x \leq 0$	F(x) = 0
$0 < x \leq 1$	F(x) = 0.15
$1 < x \leq 2$	F(x) = 0.52
$2 < x \leq 3$	F(x) = 0.80
$3 < x \leq 4$	F(x) = 0.91
$x > 4$	F(x) = 1

Il grafico di tale funzione è dunque il seguente:



La funzione di distribuzione cumulativa di una variabile discreta qualsiasi è sempre una funzione discontinua a gradini i cui salti sono localizzati nei punti corrispondenti ai valori possibili di questa variabile e sono uguali alle probabilità di questi valori. La somma di tutti i salti della funzione, in accordo con il terzo assioma della probabilità, è pari a uno.
A mano a mano che aumenta il numero di valori possibili della variabile aleatoria e diminuiscono gli intervalli tra di essi, il numero di salti diventa sempre più grande e i salti stessi più piccoli. La curva inizialmente a gradini della funzione si avvicina così a una funzione continua, caratteristica delle variabili aleatorie continue.

DISTRIBUZIONI CONTINUE

Vediamo ora in dettaglio le due funzioni fondamentali che abbiamo introdotto: la funzione di distribuzione cumulativa e la funzione densità di probabilità.

Consideriamo una variabile aleatoria continua *X* con densità di probabilità *p(x)* e il dominio elementare *dx* adiacente al punto *x*. La probabilità che tale variabile appartenga a questo intervallo è

pari a *p(x)dx*: la quantità *p(x)dx* si chiama **elemento di probabilità** (zona tratteggiata nel Grafico 1).

Se ora vogliamo esprimere la probabilità che la variabile aleatoria appartenga all'intervallo (α, β) in funzione della densità di probabilità, otteniamo

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx$$

che rappresenta la somma di tutti gli elementi di probabilità sull'intervallo (α, β).
Dal punto di vista geometrico la probabilità che la variabile aleatoria *X* appartenga all'intervallo (α, β) è uguale all'area compresa tra la curva definita della densità di probabilità (detta curva di densità) e l'asse delle ascisse, limitata dagli estremi dell'intervallo (zona tratteggiata nel Grafico 2).

Grafico 1

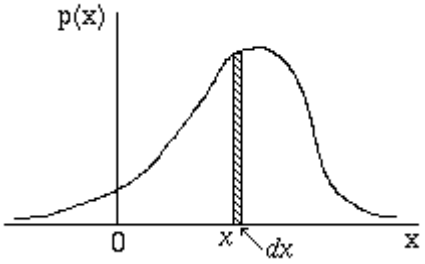
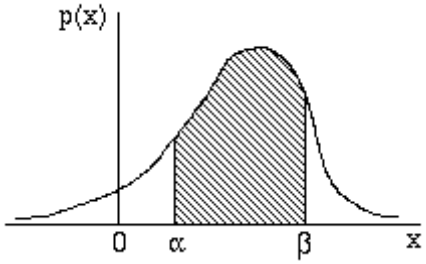


Grafico 2



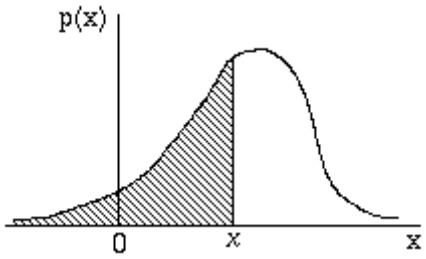
Vediamo ora come come trovare la funzione di distribuzione cumulativa partendo dalla densità di probabilità. Per definizione abbiamo:

$$F(X) = P(X < x) = P(-\infty < X < x)$$

Osservando come abbiamo ricavato la probabilità che una variabile aleatoria appartenga ad un certo intervallo o attraverso il processo inverso di derivazione della densità di probabilità, otteniamo:

$$F(x) = \int_{-\infty}^{x} p(x)\,dx$$

Geometricamente *F(x)* non è altro che l'area compresa tra la curva di densità e l'asse delle ascisse e situata a sinistra del punto *x* (zona tratteggiata).



Funzione densità di probabilità

Definita la funzione di distribuzione cumulativa, vediamo cosa succede se consideriamo la probabilità che la variabile assuma un qualsiasi valore entro un intervallo di estremi x_1 e x_2: avremo allora

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

Infatti, l'appartenenza della variabile *X* all'intervallo di estremi x_1 e x_2 equivale al verificarsi della disuguaglianza

$$x_1 \leq X < x_2$$

Esprimendo la probabilità di questo evento attraverso i seguenti tre eventi

evento *A* corrispondente a $X < x_2$
evento *B* corrispondente a $X < x_1$
evento *C* corrispondente a $x_1 \leq X < x_2$

abbiamo che l'evento *A* si può esprimere come la somma degli altri due, cioè *A* = *B* + *C*.
Per il teorema di addizione delle probabilità abbiamo

$$P(X < x_2) = P(X < x_1) + P(x_1 \leq X < x_2)$$

da cui ricaviamo la formula che avevamo introdotto

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

Operando un processo al limite per cui $\Delta x \to 0$, il rapporto tra la differenza della funzione di distribuzione cumulativa e l'intervallo stesso è la derivata

$$\lim_{\Delta x \to 0} \left[\frac{F(x + \Delta x) - F(x)}{\Delta x} \right] = \frac{dF(x)}{dx}$$

Si definisce la **densità di probabilità** o **funzione di distribuzione**:

$$p(x) = \frac{dF(x)}{dx}$$

La funzione **p(x)** caratterizza la *densità di distribuzione* dei valori della variabile aleatoria in un dato punto **x**: la densità di probabilità è una delle forme per esprimere la legge di distribuzione delle variabili aleatorie.

Condizione di normalizzazione

La condizione di normalizzazione non è altro che la generalizzazione al caso continuo del terzo assioma della probabilità.
Analogamente al caso discreto, nel caso continuo deve valere la seguente

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

che deriva dal fatto che **F(+∞)=1**.

DISTRIBUZIONI CONTINUE

Per quanto riguarda le distribuzioni di variabili aleatorie continue, noi considereremo le seguenti:

- distribuzione uniforme
- distribuzione di Gauss o normale
- distribuzione di Cauchy
- distribuzione di χ^2
- distribuzione di Student

DISTRIBUZIONI CONTINUE
- Distribuzione uniforme -

Variabili aleatorie continue di cui è noto a priori che i loro valori possibili appartengono ad un dato intervallo e all'interno di questo intervallo tutti i valori sono equiprobabili, si dicono *distribuite uniformemente* o che seguono la *distribuzione uniforme*.

Consideriamo una variabile aleatoria *X* soggetta ad una legge di distribuzione uniforme nell'intervallo (α, β) e scriviamo l'espressione della densità di probabilità *p(x)*. Questa funzione dovrà essere costante ed uguale, diciamo, a c nell'intervallo (α, β) e nulla al di fuori di esso, per cui

$$p(x) = \begin{cases} c & \text{per } \alpha < x < \beta \\ 0 & \text{altrove} \end{cases}$$

Poiché per la condizione di normalizzazione, l'area delimitata dalla curva di distribuzione deve essere uguale all'unità si ha

$$c(\beta - \alpha) = 1$$

da cui risulta

$$c = \frac{1}{\beta - \alpha}$$

La densità di probabilità *p(x)* assume quindi la forma

$$p(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{per } \alpha < x < \beta \\ 0 & \text{altrove} \end{cases}$$

Caratteristiche della distribuzione uniforme

Determiniamo le caratteristiche numeriche fondamentali della variabile aleatoria *X* soggetta alla legge di distribuzione uniforme nell'intervallo (α, β).

Il valor medio, applicando la definizione di momento iniziale di ordine 1 nel caso di variabili continue, vale

$$\overline{m} = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx = \frac{\alpha + \beta}{2}$$

In virtù della simmetria della distribuzione il valore della mediana coincide con quello del valor medio. Si tenga inoltre presente che la distribuzione uniforme non ha moda.

ora, attraverso il momento centrale di ordine 2 nel caso continuo, calcoliamo la varianza e da questa la deviazione standard:

$$D = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \left(x - \frac{\alpha + \beta}{2} \right)^2 dx = \frac{(\beta - \alpha)^2}{12}$$

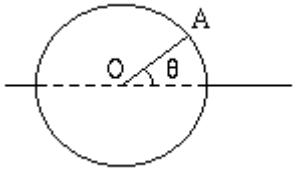
da cui la deviazione standard

$$\sigma = \sqrt{D} = \frac{\beta - \alpha}{2\sqrt{3}}$$

Esempi

- Esempio 1. Un corpo viene pesato con una bilancia di precisione avendo a disposizione solo pesi multipli di 1 g.; i risultati mostrano che il peso del corpo è compreso tra *k* e *k+1* grammi. Se si assume che questo peso sia uguale a (*k+1/2*) grammi, l'errore commesso (*X*) è, evidentemente, una variabile aleatoria, distribuita con densità di probabilità uniforme nell'intervallo (-1/2, 1/2) g.

- Esempio 2. Una ruota simmetrica in rotazione si arresta a causa dell'attrito. L'angolo formato da un raggio mobile della ruota con un raggio fissato *OA* dopo l'arresto della ruota è una variabile la cui densità di distribuzione è uniforme in $(0, 2\pi)$.



Caratteristiche della distribuzione uniforme

Determiniamo le caratteristiche numeriche fondamentali della variabile aleatoria *X* soggetta alla legge di distribuzione uniforme nell'intervallo (α, β).

Il valor medio, applicando la definizione di momento iniziale di ordine 1 nel caso di variabili continue, vale

$$\overline{m} = \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx = \frac{\alpha + \beta}{2}$$

In virtù della simmetria della distribuzione il valore della mediana coincide con quello del valor medio. Si tenga inoltre presente che la distribuzione uniforme non ha moda.

ora, attraverso il momento centrale di ordine 2 nel caso continuo, calcoliamo la varianza e da questa la deviazione standard:

$$D = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \left(x - \frac{\alpha + \beta}{2} \right)^2 dx = \frac{(\beta - \alpha)^2}{12}$$

da cui la deviazione standard

$$\sigma = \sqrt{D} = \frac{\beta - \alpha}{2\sqrt{3}}$$

LA DISTRIBUZIONE DI GAUSS
- Introduzione -

Nella teoria delle probabilità la legge di distribuzione di Gauss riveste un ruolo abbastanza importante: essa costituisce una legge limite, cui tende la maggior parte delle altre distribuzioni sotto condizioni che si verificano abbastanza di frequente.
Essa si manifesta ogni volta che la grandezza aleatoria osservata è il risultato della somma di un numero sufficientemente grande di variabili aleatorie indipendenti (o al limite debolmente indipendenti) che obbediscono a leggi di distribuzione diverse (sotto deboli restrizioni). La grandezza osservata si distribuisce seguendo la legge di distribuzione normale in un modo tanto più preciso, quanto è maggiore il numero variabili da sommare.
Numerose variabili aleatorie di uso comune, quali ad esempio gli errori di misura, si possono rappresentare come somma relativamente grande di singoli termini (errori elementari), ciascuno dei quali è dovuto ad una causa, non dipendente dalle altre. Quali che siano le leggi di distribuzione cui obbediscono gli errori elementari, le peculiarità di queste distribuzioni non si manifestano nella somma di un gran numero di termini e la somma segue una legge vicina alla legge normale.
La sola restrizione imposta agli errori da sommare è di influire relativamente poco sulla somma. Se questa condizione non si verifica e, per esempio, uno degli errori aleatori prevale nettamente sugli altri la legge di distribuzione dell'errore prevalente determina grosso modo la legge di distribuzione della somma.

Quanto enunciato fin'ora va sotto il nome di **teorema centrale del limite**.
Il teorema del limite centrale si può utilizzare anche nel caso della somma di numero relativamente piccolo di variabili aleatorie indipendenti. L'esperienza mostra che, per un numero di addendi dell'ordine di una decina, la legge di distribuzione della somma può essere approssimata con la legge normale.

DISTRIBUZIONI CONTINUE
- La distribuzione di Gauss -

La legge di distribuzione normale è caratterizzata da una densità di probabilità della forma:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

La curva descritta da tale funzione ha una forma caratteristica "a campana": tale curva è centrata sul punto di ascissa $x=m$ e in corrispondenza di esso ha il suo massimo in ordinata pari a $\frac{1}{\sigma\sqrt{2\pi}}$.
Il parametro σ è correlato alla larghezza della "campana" e, in particolare rappresenta la distanza tra l'asse di simmetria e i punti di flesso della distribuzione. Se σ è piccolo, la curva è stretta, se σ è grande, la curva è larga e più "dispersa" rispetto al valor medio m.
In particolare il parametro σ non è altro che la deviazione standard della distribuzione, così come m rappresenta il valor medio della funzione.

Solitamente la distribuzione normale viene indicata con $N(m,\sigma)$ dove m e σ sono i due parametri che caratterizzano la distribuzione.
Molto spesso si usa anche la distribuzione cosiddetta "standardizzata" $N(0,1)$ avente media 0 (centrata sull'origine degli assi) e deviazione standard pari ad 1.

DISTRIBUZIONE DI GAUSS
- Parametri fondamentali -

Calcoliamo i parametri fondamentali della legge normale e in particolare:

- la media
- la deviazione standard
- coefficienti di asimmetria e di appiattimento

Nei paragrafi seguenti si farà un discreto uso del calcolo integrale, si consiglia perciò di visitare la seguente tabella prima di proseguire.

Integrali fondamentali

La media

Utilizzando la definizione del momento iniziale di ordine 1 calcoliamo il valor medio della distribuzione di Gauss:

$$\overline{m} = \int_{-\infty}^{+\infty} x\, p(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx$$

Operando il cambio di variabile:

$$\frac{x - m}{\sigma\sqrt{2}} = t$$

otteniamo

$$\overline{m} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} (\sigma\sqrt{2}t + m) e^{-t^2}\, dt =$$

$$= \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t e^{-t^2}\, dt + \frac{m}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2}\, dt$$

Nell'ultima espressione, il primo integrale è nullo, mentre il secondo è l'integrale di Eulero-Poisson:

$$\int_{-\infty}^{+\infty} e^{-t^2}\, dt = 2 \int_{0}^{+\infty} e^{-t^2}\, dt = \sqrt{\pi}$$

Di conseguenza otteniamo che $\overline{m}=m$, ossia il parametro m rappresenta proprio il valor medio della distribuzione.

 Approfondimento

---

La deviazione standard

Per il calcolo della deviazione standard della normale sfruttiamo la definizione di varianza:

$$D = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Effettuando il cambio di variabile

$$\frac{x-m}{\sigma\sqrt{2}} = t$$

si ha

$$D = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^2 e^{-t^2} dt$$

Integrando per parti, si giunge a:

$$D = \frac{\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t \cdot 2te^{-t^2} dt = \frac{\sigma^2}{\sqrt{\pi}} \left[-te^{-t^2} \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-t^2} dt \right]$$

Il primo termine tra le parentesi quadre è nullo in quanto l'esponenziale negativo a più e meno infinito decresce molto più velocemente di quanto cresca qualsiasi potenza di *t*, mentre il secondo termine è l'integrale di Eulero-Poisson, per cui:

$$D = \sigma^2$$

Quindi il parametro σ nella formula della distribuzione normale è non è altro che la deviazione standard della distribuzione stessa.

Per quanto riguarda il parametro σ si noti che, essendo l'ordinata massima inversamente proporzionale a σ, al crescere di σ il massimo della funzione cala e la "campana" si allarga in quanto l'area delimitata dalla curva deve essere sempre uguale all'unità: al contrario, al decrescere di σ la curva di distribuzione si allunga verso l'alto restringendosi contemporaneamente ai lati, diventanto cioè sempre più piccata sul valor medio.

I coefficienti di asimmetria e di appiattimento

Prima di passare al calcolo dei coefficienti di asimmetria e di appiattimento attraverso il momento centrale di ordine 3 e quello di ordine 4, ricaviamo le formule generali per i momenti centrali di qualsiasi ordine per la distribuzione di Gauss.
Per definizione

$$\mu_s = \int_{-\infty}^{+\infty} (x-m)^s \, p(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-m)^s e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Dopo il cambio di variabile

$$\frac{x-m}{\sigma\sqrt{2}} = t$$

si ha

$$\mu_s = \frac{(\sigma\sqrt{2})^s}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^s e^{-t^2} dt$$

Integrando quest'ultima espressione per parti, si ottiene

$$\mu_s = \frac{(\sigma\sqrt{2})^2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^{s-1} \cdot t e^{-t^2} dt =$$
$$= \frac{(\sigma\sqrt{2})^2}{\sqrt{\pi}} \left[-\frac{1}{2} t^{s-1} e^{-t^2} \Big|_{-\infty}^{+\infty} + \frac{s-1}{2} \int_{-\infty}^{+\infty} t^{s-2} e^{-t^2} dt \right]$$

Tenendo conto che il primo termine entro le parentesi è nullo, rimane

$$\mu_s = \frac{(s-1)(\sigma\sqrt{2})^s}{2\sqrt{\pi}} \int_{-\infty}^{+\infty} t^{s-2} e^{-t^2} dt$$

Se con lo stesso procedimento ricaviamo il momento centrale *s-2*-esimo otteniamo

$$\mu_s = \frac{(\sigma\sqrt{2})^{s-2}}{\sqrt{\pi}} \int_{-\infty}^{+\infty} t^{s-2} e^{-t^2} dt$$

Confrontando i secondi termini delle ultime due espressioni, si vede che esse differiscono solo per il fattore $(s\text{-}1)\boldsymbol{\sigma}^2$: di conseguenza possiamo affermare che vale

$$\mu_s = (s-1)\sigma^2 \mu_{s-2}$$

Quest'ultima è una formula ricorrente che permette di esprimere i momenti di ordine superiore in termini di quelli inferiori. Tenendo conto che il momento centrale di ordine 0 ($\boldsymbol{\mu_0}$) è uguale ad 1 e quello di ordine 1 ($\boldsymbol{\mu_1}$) è uguale a 0, si possono calcolare i momenti centrali di ordine qualsiasi.
In particolare, siccome $\boldsymbol{\mu_1}$=0, segue che tutti i momenti di ordine dispari della legge normale sono nulli, mentre per quanto riguarda i momenti di ordine pari abbiamo

$\boldsymbol{\mu}_2=\boldsymbol{\sigma}^2$ $\boldsymbol{\mu}_4=3\boldsymbol{\sigma}^4$ $\boldsymbol{\mu}_6=15\boldsymbol{\sigma}^6$

In generale la formula per il momento di ordine *s* qualunque è data da

$$\mu_s = (s-1)!!\,\sigma^s$$

dove, con il simbolo *(s-1)!!* si intende il prodotto di tutti i numeri dispari tra 1 e s-1.
Poiché per la legge normale μ_3=0, il suo coefficiente di asimmetria è nullo:

$$S = \frac{\mu_3}{\sigma^3} = 0$$

in accordo con la simmetria della distribuzione di Gauss.
Dall'espressione del momento di ordine quattro ricaviamo il coefficiente di appiattimento

$$E = \frac{\mu_4}{\sigma^4} - 3 = 0$$

Ciò è del tutto naturale, in quanto il coefficiente di appiattimento caratterizza il grado di altezza raggiunto da una distribuzione effettiva in relazione alla distribuzione normale.

DISTRIBUZIONI CONTINUE
- Distribuzione di Cauchy -

La distribuzione di Cauchy, tipicamente usata da fisici, chimici, ecc. nei casi in cui si ha a che fare con fenomeni di risonanza nell'intervallo (-∞, +∞), ha la seguente forma:

$$p(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

Questa distribuzione presenta una caratteristica particolare: il suo valor medio non è definito. Infatti, applicando la definizione di momento iniziale di ordine 1, si ha:

$$\overline{m} = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x}{1+x^2}dx = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{2x}{1+x^2}dx = \frac{1}{2\pi}\ln(1+x^2)\Big|_{-\infty}^{+\infty} = \infty - \infty$$

Tuttavia, data la simmetria della distribuzione rispetto allo 0, si parla convenzionalmente di *valore principale* La varianza, calcolata come momento centrale di ordine 2, risulta infinita:

$$D = \sigma^2 = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x^2}{1+x^2}dx = \frac{1}{\pi}\left[\int_{-\infty}^{+\infty}\frac{x^2+1}{1+x^2}dx - \int_{-\infty}^{+\infty}\frac{1}{1+x^2}dx\right] =$$
$$= \frac{1}{\pi}\left[\int_{-\infty}^{+\infty}dx + \int_{-\infty}^{+\infty}\frac{1}{1+x^2}dx\right] = \frac{1}{\pi}(x - \arctan x)\Big|_{-\infty}^{+\infty} = \infty$$

Se andiamo a calcolare la dispersione tra i valori +1 e -1 otteniamo:

$$\frac{1}{\pi}\int_{-1}^{+1}\frac{1}{1+x^2}dx = \frac{1}{\pi}(\arctan x)\Big|_{-1}^{+1} = \frac{1}{\pi}\frac{\pi}{2} = \frac{1}{2}$$

Ossia si ha il 50% di probabilità di ottenere un valore compreso tra -1 e +1.

DISTRIBUZIONI CONTINUE
- Distribuzione χ^2 -

Si dice distribuzione χ^2 con ***n*** gradi di libertà la distribuzione della somma dei quadrati di ***n*** variabili aleatorie indipendenti, ciascuna delle quali obbedisce ad una legge normale con speranza matematica nulla e varianza pari all'unità.
Considerata cioè la variabile

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

si ha che la nuova variabile χ^2 con definita come

$$\chi^2 = \sum_1^n x_i^2$$

si distribuisce secondo quella che viene chiamata distribuzione χ^2 per ***n*** gradi di libertà, ossia

$$p(\chi^2) = \frac{1}{(2)^{\frac{n}{2}}\,\Gamma\left(\frac{n}{2}\right)} (\chi^2)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}$$

Il numero *n* delle variabili addende indipendenti rappresenta il *numero di gradi di libertà* e tale numero costituisce il **valor medio** della distribuzione in quanto, essendo le singole variabili x_i distribuite secondo una normale standardizzata, i valori medi delle x_i^2 sono tutti uguali ad 1.

$$\overline{x^2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = D = 1$$

Infatti calcolare il valor medio di x_i^2 equivale a calcolare la varianza ***D*** di x_i che è appunto uguale ad 1 poiché le variabili sono standardizzate.
La **varianza** della distribuzione χ^2, invece, è uguale a ***2n***.

Esempio

Consideriamo una variabile aleatoria ***X*** distribuita normalmente: siano x_i le sue realizzazioni in ***n*** prove.
Introduciamo la seguente variabile aleatoria

$$v = \frac{(n-1)\tilde{D}}{D}$$

dove $\tilde{D}$ è il quadrato della deviazione standard nella sua forma "migliorata" calcolata attraverso le ***n*** realizzazioni della variabile ***X***

$$\tilde{D} = \frac{\sum_{i=1}^{n}(x_i - \bar{m})^2}{n-1}$$

mentre ***D*** è la varianza della distribuzione normale.
Questa nuova variabile obbedisce alla distribuzione χ^2.

Vediamo ora un esempio di test del χ^2. consideriamo un esperimento in cui la distribuzione di Poisson è la distribuzione attesa. Supponiamo di allestire un contatore Geiger per contare le particelle di raggi cosmici che arrivano in una certa regione e inoltre di contare il numero di particelle che arrivano in 100 intervalli separati di un minuto con i risultati riportati in tabelle

Conteggi α	Occorrenza	Intervallo k	Osservazioni O_k	Numero atteso T_k
0	7	1	7	7.5
1	17	2	17	19.4
2	29	3	29	25.2
3	20	4	20	21.7
4	16	5	16	14.1
5	8	6	11	12.1
6	1			
7	2			
8 o più	0			

L'osservazione delle occorrenze in tabella suggerisce di raggruppare tutti i conteggi maggiori di 5 in un singolo intervallo.

L'ipotesi che vogliamo verificare è che il numero di conteggi è governato da una distribuzione di Poisson. Poiché il conteggio medio atteso è incognito, dobbiamo prima calcolare la media dei nostri 100 conteggi.

DISTRIBUZIONI CONTINUE
- Distribuzione di Student -

Per ricavare la distribuzione di Student consideriamo una variabile aleatoria ***y*** distribuita secondo una distribuzione normale standard, e una variabile ***x***, indipendente da ***y***, avente una distribuzione χ^2 con *n* gradi di libertà.
Siamo allora interessati a studiare il comportamento della nuova variabile *t* definita come:

$$t = \sqrt{n}\,\frac{y}{\sqrt{x}}$$

Questa nuova variabile obbedisce alla *distribuzione di Student*

$$S_n(t) = \frac{1}{\sqrt{\pi n}}\,\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

dove la funzione **Γ** è così definita:

$$\Gamma(n) = \int_0^{\infty} u^{n-1} e^{-u} du$$

Questa distribuzione presenta due comportamenti differenti a seconda del valore di ***n***: per piccoli valori di ***n*** essa approssima la distribuzione di Cauchy, mentre per ***n*** grandi approssima la normale. Generalmente questa distribuzione viene adoperata quando il numero di valori considerati è piccolo (***n*** piccolo), quando si studiano eventi molto distanti dal valor medio o quando si verificano entrambe le situazioni.

Infine si noti che gli unici momenti finiti della distribuzione di Student sono quelli di ordine minore di ***n***. In particolare tutti i momenti di ordine dispari sono nulli in virtù della simmetria della distribuzione, mentre tutti quelli di ordine pari superiore ad ***n*** sono infiniti.

Esempio

Consideriamo una variabile aleatoria distribuita normalmente: siano x_i le sue realizzazioni in ***n*** prove.
Introduciamo la seguente variabile aleatoria

$$t = \sqrt{n}\,\frac{\tilde{m} - m}{\sqrt{\tilde{D}}}$$

dove $\tilde{m}$ è la media calcolata sugli ***n*** valori ottenuti

$$\tilde{m} = \frac{\sum_{i=1}^{n} x_i}{n}$$

e $\tilde{D}$ è il quadrato della deviazione standard nella sua forma "migliorata"

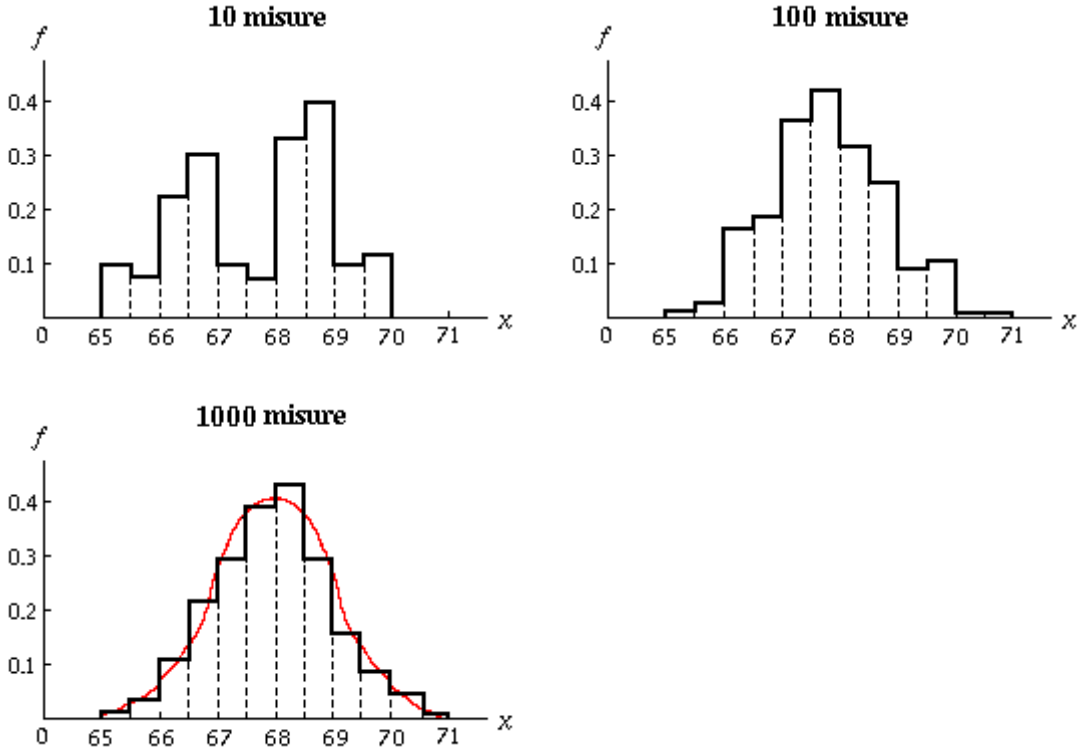
$$\tilde{D} = \frac{\sum_{i=1}^{n} (x_i - \tilde{m})^2}{n-1}$$

La nuova variabile ***t*** così definita obbedisce alla distribuzione di Student.

DISTRIBUZIONI LIMITE

Si parla di distribuzioni limite nel momento in cui il numero di eventi considerati si avvicina all'infinito.
Con la dicitura "si avvicina all'infinito" si intende un numero di eventi, o realizzazioni di una variabile aleatoria, sufficientemente grande, ma anche qui la nostra definizione, di fatto, è ancora poco chiara.
Quello che si vuole dire, quando si parla di distribuzioni limite, è che dopo un certo numero di eventi, i risultati ottenuti si disporranno secondo una determinata distribuzione, mostreranno cioè, al passare delle prove, una determinata distribuzione. Tale distribuzione diviene sempre più evidente all'aumentare degli eventi e, al limite per il numero di realizzazioni che tende all'infinito, la distribuzione assume un carattere continuo.

Vediamo di illustrare quanto detto con un esempio. Consideriamo i tre istogrammi sottostanti i quali mostrano le frazioni (*f*) di misure che cadono nei rispettivi intervalli (*x=cm*) nei casi in cui siano state effettuate 10, 100 e 1000 misure della stessa grandezza.



Si vede abbastanza bene che all'aumentare del numero di misure che effettuiamo la distribuzione dei dati ottenuti si avvicina ad una curva continua, definita. Quando ciò si verifica, la curva continua è detta *distribuzione limite*.
La distribuzione limite è una costruzione teorica, che non può mai essere misurata esattamente: più misure facciamo e più il nostro istogramma si avvicina alla distribuzione limite. Ma soltanto se noi fossimo in grado di effettuare un numero infinito di misure ed di utilizzare intervalli infinitamente stretti, allora otterremmo realmente la distribuzione limite stessa.

SISTEMI DI DUE VARIABILI ALEATORIE
- Caratteristiche numeriche -

Nelle applicazioni pratiche della teoria delle probabilità occorre spesso risolvere problemi in cui i risultati delle prove vengono descritti da due o più variabili aleatorie che formano un *insieme* o *sistema*.
Per esempio il punto di impatto di un proiettile viene determinato da due grandezze: l'ascissa e l'ordinata, che si possono definire come un sistema di due variabili aleatorie.
Analogamente, il punto di scoppio del proiettile viene considerato come un sistema di tre variabili.
Estendo l'esempio bellico in questione, quando vengono sparati *n* colpi, le coordinate dei punti di impatto sul piano del bersaglio costituiscono un insieme di *2n* variabili aleatorie.

Senza andare ad introdurre le funzioni che caratterizzano tali sistemi (che richiederebbero inoltre la conoscenza di integrali multidimensionali), introduciamo due concetti che ci saranno utili e sono:

covarianza
coefficiente di correlazione

SISTEMI DI DUE VARIABILI ALEATORIE
- La covarianza -

Per definire il concetto di covarianza per due variabili aleatorie, dobbiamo prima introdurre la generalizzazione del concetto di momento. Per un sistema di due variabili aleatorie discrete (*x*,*y*) si dice **momento iniziale di ordine k,s** la seguente espressione:

$$\alpha_{k,s} = \sum_i \sum_j x_i^k y_j^s \, p_{ij}$$

che non è altro che la speranza matematica del prodotto delle potenze *k* e *s* delle due variabili e dove p_{ij} è la probabilità che il sistema definito dalle due variabili assuma il valore (x_i,y_j). Analogamente possiamo definire il **momento centrale di ordine k,s** nel modo seguente:

$$\mu_{k,s} = \sum_i \sum_j (x_i - \overline{x})^k (y_j - \overline{y})^s \, p_{ij}$$

Ovviamente le generalizzazioni al caso continuo saranno

$$\alpha_{k,s} = \int_{-\infty}^{+\infty} \!\!\int x^k y^s \, f(x,y) \, dx \, dy$$

e

$$\mu_{k,s} = \int_{-\infty}^{+\infty} \!\!\int (x - \overline{x})^k (y - \overline{y})^s \, f(x,y) \, dx \, dy$$

dove *f(x,y)* è la densità di probabilità del sistema.

Il *momento centrale misto del secondo ordine* $K_{x,y}$ ha un'importanza particolare:

$$\mu_{x,y} = K_{x,y} = \sum_i \sum_j (x_i - \overline{x})(y_j - \overline{y}) \, p_{ij}$$

o, nel caso continuo

$$\mu_{x,y} = K_{x,y} = \int_{-\infty}^{+\infty} \iint (x - \overline{x})(y - \overline{y})\, f(x,y)\, dx\, dy$$

Tale momento è detto **covarianza**. E` un parametro che, oltre descrivere la dispersione delle variabili, esprime anche la loro *relazione*. In particolare la covarianza di variabili aleatorie **indipendenti** è **nulla**, mentre è diversa da zero se esiste una relazione che lega le variabili considerate. Ma la covarianza caratterizza non solo la dipendenza delle variabili aleatorie, bensì anche la loro dispersione. Infatti nel caso in cui una delle variabili del sistema differisca di poco dal suo valore atteso, la covarianza sarà piccola qualunque sia il grado di legame tra le due variabili.

Dunque per caratterizzare solo il legame tra le variabili si passa dal momento $K_{x,y}$ ad un'altra grandezza adimensionale detta coefficiente di correlazione.

SISTEMI DI DUE VARIABILI ALEATORIE
- Il coefficiente di correlazione -

Per caratterizzare il legame tra due variabili aleatorie (*x*,*y*) si ricorre alla grandezza adimensionale

$$r_{x,y} = \frac{K_{x,y}}{\sigma_x \sigma_y}$$

dove $K_{x,y}$ è la covarianza e $\sigma_{<I<x}$ e σ_y sono gli scarti quadratici medi delle due variabili. E` evidente che il coefficiente di correlazione si annulla mutuamente con la covarianza: di conseguenza il coefficiente di correlazione delle variabili indipendenti è nullo.
Bisogna altresì considerare che non sempre variabili non correlate sono indipendenti: questo dipende dal fatto che il coefficiente di correlazione evidenzia solamente una correlazione di tipo lineare. il coefficiente di correlazione è quindi una misura del della forza della relazione lineare fra le due variabili.
Se le variabili aleatorie (*x*,*y*) sono legate dalla relazione funzionale esatta

$y=ax+b$

allora $r_{x,y}$ è uguale a +1 o -1; nel caso generale il valore del coefficiente è compreso nei limiti

$-1<R_{x,y}<+1$

Per $r_{x,y}> 0$ la correlazione è detta *positiva*, mentre per $r_{x,y}< 0$ essa si dice *negativa*. Una correlazione positiva significa che al crescere di una variabile, l'altra tende ugualmente a crescere in media; una correlazione negativa significa che al crescere di una variabile, l'altra tende a decrescere in media.

NUMERI CASUALI

SEQUENZE CASUALI

Cosa significa dire che una sequenza di numeri è casuale? Consideriamo le due sequenze formate da 0 e 1:

a) **1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0** b) **1 0 0 1 1 0 1 1 0 1 1 1 0 0 1 0 0 0 0 1**

Intuitivamente siamo portati ad affermare che la sequenza **a)** non è casuale poiché è riconoscibile una regolarità o, detto in un altro modo, un algoritmo in grado di generarla, del tipo "*scrivi dieci volte 1 0*".
Al contrario, per la sequenza **b)**, non siamo in grado di discernere alcuno schema soggiacente né predire lo sviluppo successivo. Non sembra sussistere alcuna regola che presieda alla formazione della sequenza e, quindi, attribuiamo la sua generazione ad un evento casuale. Per esempio, possiamo pensare di aver lanciato in aria per venti volte una moneta non truccata e aver trascritto 0 ad ogni uscita testa e 1 ad ogni uscita croce.
A simili argomentazioni , però si può controbattere osservando che nulla vieta di ritenere che anche la sequenza a) possa essere stata generata associando 0 e 1 ai risultati dei lanci della stessa moneta che, in tal caso, si sarebbero manifestati con testa e croce alternativamente per venti volte.

Consideriamo altre due sequenze formate dai primi dieci numeri interi:

c) **0 1 2 3 4 5 6 7 8 9** d) **1 8 1 3 1 9 4 9 8 5**

È immediato ritenere la sequenza **c)** non casuale e casuale la sequenza **d)**. Eppure, anche la sequenza, **d)** apparirà *non* casuale, qualora si conosca il metodo con cui è stata generata. In particolare, si è scritto in lettere la sequenza **c)** e poi si sono sommate le cifre corrispondenti, nell'alfabeto italiano, alle lettere con cui si sono scritti i numeri precedenti: infine si sono sommate le cifre che formano la somma precedente e, nel caso che il risultato superi 9, si è eseguito una somma ulteriore.
Pertanto le due sequenza **c)** e **d)** sono concettualmente equivalenti: l'unica distinzione risiede nella diversità dell'algoritmo di generazione.

È dunque la casualità un concetto *soggettivo*, basato sul fatto di sapere o non sapere come la sequenza è stata prodotta?
Se così fosse quello che è casuale per una persona all'oscuro del metodo di generazione non lo è per altri che ne sono a conoscenza. Esiste tuttavia una via per definire *oggettivamente* la casualità di una sequenza di numeri?

SEQUENZE CASUALI
- Metodi di generazione -

Il fatto che si possano produrre solo numeri casuali che, proprio per il fatto di essere generati attraverso un algoritmo, casuali non sono, appare in un qualche modo paradossale.
La verità è che noi li riteniamo casuali fintanto che rimaniamo all'oscuro delle "regole del gioco" con cui sono stati ottenuti: regole che possono essere scoperte, qualora si sottoponga la sequenza ad opportuni test statistici.
Sembrerebbe conveniente, pertanto, ricorrere a fenomeni fisici intrinsecamente non deterministici per generare sequenze di eventi casuali, da usarsi direttamente oppure per produrre tabelle da cui estrarre ogni volta che si desidera successioni di numeri a caso.

Distinguiamo i metodi di generazione in:

metodi manuali estrazione da tabelle generazione con il calcolatore

Un'ulteriore approfondimento viene riservato alla generazione con il calcolatore, essendo questo il metodo ampiamente più diffuso al giorno d'oggi.

GENERAZIONE CON IL CALCOLATORE
- Metodi congruenziali -

Oggi i metodi di generazione maggiormente usati sono i cosiddetti *metodi congruenziali*, così chiamati perché si basano sulla nozione di congruenza.
Dati tre numeri interi *x*, *y*, *z* si dice che "*x è congruo ad y modulo z*" e si scrive

$$x = y \pmod{z}$$

se *x* è la parte intera del resto della divisione tra *y* e *z*.

Per generare numeri casuali, si distinguono tre tipi di metodi congruenziali:

1. Metodo additivo:

$$r_{i+1}=r_i+r_{i-1} \pmod{m}$$

2. Metodo moltiplicativo:

$$r_{i+1}=ar_i \pmod{m}$$

3. Metodo misto:

$$r_{i+1}=(ar_i+b) \pmod{m}$$

dove r_i, a, b, m sono interi *non negativi* e r_i è compreso tra 0 ed m.
Vengono chiamati: a *moltiplicatore*, b *incremento* e m *modulo*, mentre r_o, valore iniziale della sequenza, è detto *seme*.
Tutti questi metodi danno origine a *sequenze periodiche*, con periodo minore o al più uguale ad m.

Se si desiderano numeri casuali razionali nell'intervallo **[0,1)**, questi si ottengono facilmente mediante la divisione:

$$u_i = r_i / m$$

Per ciò che riguarda il periodo, si cerca sempre di scegliere m in modo tale da assicurarsi una sequenza abbastanza lunga; in ogni caso, la lunghezza del periodo deve essere superiore alla quantità di numeri casuali richiesti per la soluzione del problema che interessa.
Quando la lunghezza del periodo è uguale ad m, si dice che il generatore ha *periodo pieno*.

Oltre a dar luogo ad un periodo sufficientemente lungo, è auspicabile che il generatore possegga un'alta velocità di generazione, vale a dire che sia in grado di effettuare il calcolo dell'espressione, ad esempio, $(ar_i+b) \pmod m$ il più rapidamente possibile.

GENERATORI RANDOM
- Controlli statistici -

Ad una sequenza vengono applicati dei test per verificare se sia lecito riguardare i numeri prodotti *come se* fossero le realizzazioni della variabile casuale uniformemente distribuita nell'intervallo compreso tra 0 e 1. I principali test che si utilizzano sono:

- Test del χ^2
- Test seriale o di autocorrelazione
- Test del poker
- Test dell'up and down
- Test di Kolmogorov-Smirnov

Si noti che nella trattazione dei test useremo la notazione u_i per definire numeri casuali reali compresi tra **0** e **1**, mentre si denoteranno con t_i i numeri interi compresi tra **0** e z-**1** dove

$$t_i = z u_i$$

Nel caso in cui si abbia z=**2** e t_i=**1**, si ottiene una sequenza di 0 e 1, mentre per z=**10** si ottengono sequenze di numeri interi compresi tra 0 e 9.

CONTROLLI STATISTICI
- Test del chi-quadrato -

Il test del chi-quadrato riveste un ruolo fondamentale tra i vari tipi di controlli statistici che si possono effettuare su sequenze di numeri casuali: infatti tutti gli altri test che abbiamo elencato

(seriale, poker,...) si riconducono ad un'analisi di grandezze secondo il test chi-quadrato.
Senza esporre nuovamente la parte relativa alla teoria del test, andiamo a vedere come operativamente viene applicato su sequenze di numeri random.

Supponiamo di avere a che fare con una sequenza di numeri casuali distribuiti nell'intervallo **[0,1)**: vogliamo controllare se tali valori possono essere considerati alla stregua delle realizzazioni di una variabile casuale ***U*** uniformemente distribuita all'interno di tale intervallo.
La prima cosa da fare è dividere l'intervallo **[0,1)** in ***k*** sottointervallli, ad esempio 100:

[0.00 , 0.01) [0.01 , 0.02) [0.02 , 0.03) .....

Dopodiché generiamo *N* numeri casuali u_i, per esempio 100000. Ci aspettiamo che questi 100000 numeri si equipartiscano nei 100 sottointervalli. Naturalmente in ogni intervallo non ne cadranno esattamente ***N/k***, ma, diciamo, n_i.
Quanto vicino ad ***N/k*** deve essere ciascun valore n_i perché la sequenza sia accettata?
Seguendo la filosofia tipica del test del χ^2, si considera la quantità

$$Y = \sum_{i=1}^{k} \frac{(n_i - N/k)^2}{N/k}$$

Nel caso in cui il numero di intervalli sia abbastanza grande e lo sia anche ogni n_i, la variabile ***Y*** segue la distribuzione χ^2 con **k-1** gradi di libertà. Fissato un livello di confidenza ***a*** (di solito: 0.01 o 0.05), tavole o codici numerici forniscono il valore χ_a^2 tale che la probabilità che la variabile casuale χ^2 assuma un valore maggiore di χ_a^2 sia uguale ad ***a***.

CONTROLLI STATISTICI
- Test di autocorrelazione -

Il *test di autocorrelazione* è riconducibile ad un caso particolare del test del χ^2 utilizzando le *coppie* di numeri interi t_i compresi tra 0 e 9, ottenuti ponendo ***z*=10**.
Ci si aspetta ragionevolmente che, ad esempio, uno 0 sia seguito da un altro 0 in **1/10** dei casi e così da un 1, un 2, ecc.

Le z^2 coppie (0-0, 0-1, 0-2,..., 9-8, 9-9) che si formano svolgono così il ruolo delle categorie ***k*** per il test del χ^2. Se il segmento a cui si applica il test è lungo ***2N***, l'aspettativa teorica è che ogni coppia si presenti N/z^2 volte.
Poiché nel test del χ^2 si richiede che ***N*>5*k***, per evitare di trattare con sequenze eccessivamente lunghe, occorre diminuire il valore di ***z***, ma non oltre una certa soglia perché il test abbia ancora senso.
Se poi il test è esteso a triplette, queste diventano z^3 e ciò impone un'ulteriore diminuzione di ***z***.
Oltre le triplette il test non è quasi più proponibile.

CONTROLLI STATISTICI
- Test del poker -

Il *test del poker* consiste nel considerare, entro segmenti di sequenza lunghi ***5N***, ***N*** gruppi di 5 numeri consecutivi t_i che chiamiamo ***a, b, c, d, e***.
I gruppi vengono suddivisi in sette categorie corrispondenti ai seguenti "punteggi" del poker (nella terza colonna vengono riportate le relative probabilità):

Tutti diversi	***a, b, c, d, e***	0.3024
Coppia	***a, a, b, c, d***	0.5040
Doppia coppia	***a, a, b, b, c***	7.2×10^{-2}
Tris	***a, a, a, b, c***	7.2×10^{-2}
Full	***a, a, a, b, b***	0.9×10^{-3}
Poker	***a, a, a, a, b***	0.45×10^{-3}
Colore	***a, a, a, a, a***	0.1×10^{-3}

Si confronta tramite un test del tipo chi-quadrato la probabilità dei vari punteggi con la frequenza con cui questi si presentano nella sequenza in esame.

CONTROLLI STATISTICI
- Test dell'up and down -

Questo test mira a controllare se la sequenza in esame contiene lunghe stringhe monotone. Per esempio consideriamo la seguente stringa di t_i:

3,5,2,3,7,8,5,1

Si confronta t_i con t_{i+1} per stabilire se è maggiore o minore. Così nell'esempio si ha:

3 - 5	sequenza su lunga 1
5 - 2	sequenza giù lunga 1
2 - 3 - 7 - 8	sequenza su lunga 3
8 - 5 -1	sequenza giù lunga 2

Da un buon generatore ci si aspetta che non produca stringhe monotone troppo lunghe. Se vi sono più di *n* stringhe lunghe *d*, il test denuncia che il generatore non è accettabile.

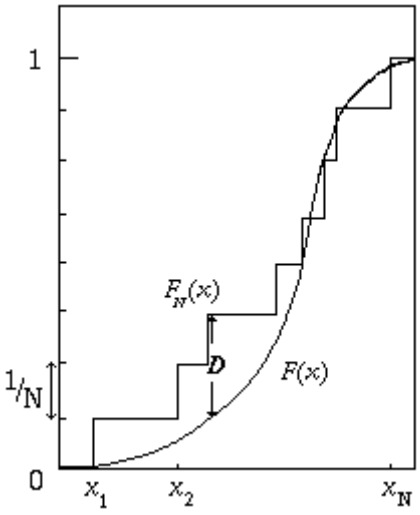
CONTROLLI STATISTICI
- Test di Kolmogorov-Smirnov -

Il test di Kolmogorov-Smirnov si applica a distribuzioni continue ponendo a confronto la distribuzione cumulativa teorica con la distribuzione cumulativa osservata.

Il test si basa sulla differenza ***D*** che rappresenta la differenza massima, in valore assoluto, tra le due distribuzioni cumulative ed è definita come segue:

$$D = \max_{-\infty < x < +\infty} |F_N(x) - F(x)|$$

dove *F(x)*è la funzione di distribuzione cumulativa teorica e $F_N(x)$ è la funzione di distribuzione cumulativa campionaria.
Se $x_1, x_2, ..., x_N$ è un campione casuale, $F_N(x)$ rappresenta il numero (rapportato ad *N*) degli elementi del campione x_i, minori od uguali ad *x*. Nella figura che segue sono rappresentate la *F(x)* e la $F_N(x)$: si vede che, da come è stata definita, $F_N(x)$ è costante tra x_i e x_{i+1} e ogni gradino è pari ad 1/*N*.



Per esempio, se il campione è costituito da una successione di u_i, con ***i*** compreso tra 1 ed *N*, si calcola il valore assoluto della differenza ***D*** tra la funzione di distribuzione cumulativa $F_N(u)$ e la funzione di distribuzione cumulativa della variabile casuale ***U***, uniformemente distribuita in (0,1). Analogamente al caso del test del chi-quadrato, ***D*** viene confrontato con certi valori critici D_a (ad esempio *a*=0.01 o *a*=0.05) per decidere se accettare o meno l'ipotesi che i numeri u_i siano equamente distribuiti.
A differenza del test del chi-quadrato, il test di Kolmogorov e Smirnov offre il vantaggio di non richiedere che i dati siano in qualche modo raggruppati e, quindi, non dipende da come vengono scelte le categorie nelle quali ripartire la sequenza.

METODO DI MONTE CARLO
- Generalità -

Quando von Neumann chiamò *"di Monte Carlo"* un procedimento matematico basato sull'utilizzazione di numeri casuali intendeva, con quel nome, alludere alla capitale del Principato di Monaco. Più precisamente si riferiva al casinò, alle sale da gioco, alle roulette; congegni questi escogitati appositamente affinché la fortuna o la rovina di coloro che per vizio, noia o altro si cimentano nel gioco, sia decretata dalla assoluta casualità che si manifesta nell'uscita di questo o quel numero.

Se vogliamo precisare esattamente cosa sia il "metodo di Monte Carlo", ci troviamo subito di fronte ad una difficoltà. Basta, infatti, una scorsa alla letteratura per rendersi conto di quanto sia controversa la definizione.
Una definizione tutto sommato soddisfacente è la seguente:

> *il metodo di Monte Carlo consiste nel cercare la soluzione di un problema, rappresentandola quale parametro di una ipotetica popolazione e nello stimare tale parametro tramite l'esame di un campione della popolazione ottenuto mediante sequenze di numeri casuali.*

Per esempio supponiamo, supponiamo che **I** sia il valore incognito da calcolare e che si possa interpretare quale valore medio di una variabile casuale **X**. Il metodo di Monte Carlo consiste, in questo caso, nello stimare **I** mediante il calcolo della media di un campione costruito determinando ***N*** valori di **X**; ciò si ottiene tramite un procedimento che prevede l'uso di numeri casuali.

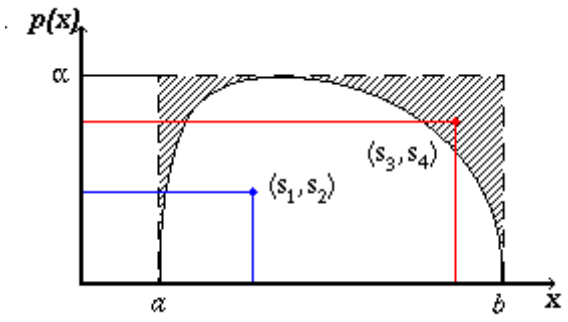
METODO DI MONTE CARLO
- Il metodo del chiama e scarta -

Il metodo del chiama e scarta (*hit or miss*) è dovuto a von Neumann e trova una delle sue più naturali applicazioni nella risoluzione degli integrali non risolvibili analiticamente: vediamo ora in dettaglio come tale metodo opera.

Sia una variabile casuale ***X*** definita nell'intervallo [a,b] e sia la funzione densità di probabilità:

$$p(x) \leq \alpha \qquad a \leq x \leq b$$

Si tratta di scegliere un punto "a caso" entro il rettangolo di base (b-a) e altezza α.



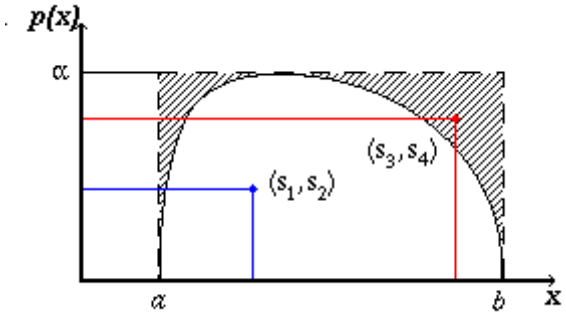
Se tale punto giace sotto la curva definita dall'equazione ***y=p(x)*** è accettato, come ad esempio il punto di coordinate (s_1,s_2), altrimenti viene scartato, come per esempio il punto di coordinate (s_3,s_4). Per ottenere i punti a caso si sfruttano i numeri casuali.

È chiaro che con questo metodo è necessario poter esprimere analiticamente la funzione densità di probabilità, ma non si richiede la conoscenza della funzione cumulativa, nè tantomeno della sua inversa.

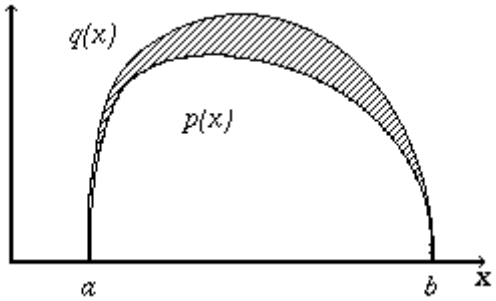
METODO DI MONTE CARLO
- Il metodo del chiama e scarta -

Vista la filosofia del metodo chiama e scarta, dovrebbe risultare altresì intuibile che il metodo è tanto più efficiente quanto più piccolo è il numero dei punti che vengono scartati. Allo scopo sono stati previsti alcuni miglioramenti in conformità della forma della ***p(x)***.

Ad esempio, invece di considerare un rettangolo siffatto



si può pensare di racchiudere la densità ***p(x)*** all'interno di una funzione ***q(x)*** tale che segua da vicino l'andamento della ***p(x)***, al fine di minimizzare le aree.



Infatti è tale rapporto che determina l'**efficienza** del metodo, piuttosto che i dettagli della forma delle due funzioni.
Poiché in questo modo occorrono le realizzazioni di una variabili casuale ***q(x)***, è opportuno scegliere quest'ultima in modo da rendere agevole il processo di generazione.

METODO MONTE CARLO
- Gli integrali -

Occupiamoci ora di come sia possibile stimare una quantità incognita attraverso il metodo statistico di Monte Carlo.
Nel nostro caso si tratterà della stima di un *integrale*. Va detto che molto spesso altri metodi numerici di risoluzione sono più consigliabili ed efficaci; ciò nondimeno, per integrali multidimensionali il Monte Carlo può risultare competitivo o, addirittura, la scelta più opportuna

quando il confine che delimita il dominio di integrazione è complicato, l'integrando non esibisce picchi concentrati in regioni molto ristrette e, infine, non si pretende un'accuratezza molto elevata.

Cominciamo con un semplice integrale:

$$I = \int_a^b f(x)dx$$

dove la ***f(x)*** è una funzione continua sull'intervallo di integrazione **[a,b]**.
Riscriviamo la precedente come:

$$I = \int_a^b f(x)\frac{p(x)}{p(x)}dx$$

dove ***p(x)*** è una qualsiasi funzione densità di probabilità definita in **[a,b]**.
Sia ***X*** una variabile casuale con densità di probabilità ***p(x)*** e sia ***G*** la variabile casuale definita da:

$$G = \frac{f(x)}{p(x)}$$

Poiché, per la definizione di valore atteso, si ha:

$$\overline{X} = \int_a^b x\ p(x)\ dx$$

nel nostro caso ne consegue che

$$\overline{G} = \int_a^b \frac{f(x)}{p(x)}p(x)\ dx = I$$

La valutazione di un integrale diventa, pertanto, un problema risolvibile in chiave statistica, stimandone il valore mediante l'esame di un opportuno campione.
Detta ***g*** una generica realizzazione della variabile aleatoria ***G***, la stima dell'integrale ***I*** diventa:

$$I = \overline{g} = \frac{1}{N}\sum_{i=1}^{N} g_i = \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)}$$

Siffatto metodo, in cui un integrale è rappresentato come un valore medio, è detto ***Monte Carlo grezzo***

BIBLIOGRAFIA

Mario Severi - *Introduzione alla sperimentazione fisica* Ed. Zanichelli - 1982
John R. Taylor - *Introduzione all'analisi degli errori* Ed. Zanichelli - 1986
Elena S. Ventsel - *Teoria delle probabilità* Ed. MIR - 1983