

DALLA TEORIA DEGLI ERRORI AL TRATTAMENTO DEI DATI

LA MISURA DELLE GRANDEZZE

Nel descrivere i fenomeni, occorre da un lato *elaborare dei modelli* (cioè delle relazioni matematiche fra le grandezze, che consentano di descrivere e prevedere il fenomeno) e dall'altro darsi degli strumenti per verificare il grado di approssimazione di queste elaborazioni (essenzialmente interrogando la realtà fisica, cioè *misurando grandezze*).

In natura ci sono delle quantità differenti tra loro, ma che, all'interno della stessa specie, possono essere poste in relazione tra loro. Si definisce grandezza tutto ciò di cui si può dire che è più o meno rispetto a qualcosa della stessa specie: un tronco più lungo di un altro, una quantità d'acqua meno rilevante di un'altra...

Una grandezza è una caratteristica che viene riconosciuta come comune in singole concretizzazioni di concetti che nascono dall'osservazione della realtà.

"...dicesi misura di una grandezza... qualsiasi metodo con cui si stabilisca una corrispondenza univoca e reciproca fra tutte le grandezze di un determinato genere e tutti i numeri interi, razionale e reali secondo il caso ... la misurazione richiede una relazione uno - uno tra il numero e le grandezze in questione: relazione che può essere diretta o indiretta... secondo le circostanze" (Bertrand Russel, I principi della matematica)

Quando si effettua la misura di una quantità di grandezza lo scopo dell'operazione è quello di associare in modo **univoco** un numero alla quantità di grandezza sottoposta all'operazione di misura. Quantità di grandezza e misure devono corrispondersi univocamente.

Ad ogni quantità di grandezza deve cioè corrispondere una ed una sola misura e ad un numero deve corrispondere, nell'ambito della stessa classe di grandezza, una ed una sola quantità di grandezza.

Una grandezza può essere misurata con:

- **osservazioni dirette**: quelle nelle quali si conta quante volte la grandezza campione è contenuta nella grandezza osservata. Un esempio di determinazione

diretta è la misurazione di una lunghezza con un nastro metrico.

- **osservazioni indirette**: quelle che permettono di ottenere la misura di una grandezza misurandone altre. Sono definite da un legame funzionale che lega una grandezza ad altre direttamente misurabili. (es: $\text{area} = l_1 \times l_2$).
- **osservazioni condizionate**: quelle che devono soddisfare ad una determinata equazione di condizione (es: somma angoli interni di un triangolo = 180°).

L'INCERTEZZA DELLA MISURA

L'esigenza di introdurre l'incertezza nasce da una osservazione sperimentale: la ripetizione della misura di una medesima grandezza in talune condizioni porta a risultati diversi, sia pure tutti raggruppati in un intervallo limitato.

Per ridurre ad un unico valore la molteplicità di numeri che si riferiscono ad una stessa quantità di grandezza, si devono cercare le cause che generano questa variabilità di risultati della misura ripetuta e definire delle modalità per ricavare un unico valore dalla molteplicità dei valori ottenuti mediante le operazioni di misura ripetute.

Tali cause vengono individuate in due possibili categorie, una legata più propriamente ai limiti imposti dagli strumenti con cui le operazioni di misura vengono effettuate, l'altra legata all'ambiente in cui tali operazioni hanno luogo.

Si definisce **sensibilità** di uno strumento la più piccola quantità di grandezza misurabile univocamente con esso.

Esempi:

- per un righello millimetrato la sensibilità è 1 mm;
- per una bilancia con scala graduata in grammi la sensibilità è 1 grammo.

Si definisce **precisione** di uno strumento il rapporto tra la sensibilità dello strumento e la massima quantità di grandezza (range) che lo strumento può misurare. La precisione è quindi un numero adimensionale; si dice che la precisione di uno strumento è tanto maggiore quanto minore è il numero che la esprime.

Esempi:

- una riga di 1 metro con suddivisione in millimetri ha una precisione di

$$\frac{1\text{mm}}{1000\text{mm}} = 10^{-3}$$

- una bilancia che può pesare una massa di entità massima di 10 kg e avente una graduazione in grammi ha una precisione di

$$\frac{1\text{g}}{10000\text{g}} = 10^{-4}$$

Per il fatto di essere adimensionale, la precisione ci permette di confrontare l'accuratezza di misure di diverso tipo che intervengono nella determinazione di una grandezza misurata indirettamente.

Esempio:

Una termometro per la misura corporea ha un range dai 34° ai 41° con graduazione in gradi ha una precisione di

$$\frac{1g}{70g} = 0.014;$$

un metro estensibile di 2m con suddivisione in mm ha una precisione di

$$\frac{1mm}{2000mm} = 0.0005;$$

si può affermare che il metro estensibile sia più preciso.

Una delle cause che crea la mancanza di univocità sui valori ottenuti nel ripetere la misura di una stessa quantità di grandezza, risiede nel fatto che generalmente noi usiamo gli strumenti pretendendo di aumentare con operazioni di stima la sensibilità, oppure con operazioni ripetitive la precisione.

Ad esempio misuriamo una lunghezza con un righello millimetrato e stimiamo i decimi di millimetro se la lunghezza non risulta uguale ad un numero finito di millimetri.

Oppure misuriamo una lunghezza di decine di metri riportando più volte una riga di un metro, commettendo delle imprecisioni.

Questi due fatti, cioè

- usare uno strumento al di fuori del suo campo di precisione,
- pretendere di aumentarne la sensibilità con un'operazione di stima,
esempio: si misura una lunghezza con un righello millimetrato e si stimano i decimi di millimetro se la lunghezza non risulta uguale ad un numero finito di millimetri. Oppure si misura una lunghezza di decine di metri riportando più volte una riga di un metro, commettendo delle imprecisioni.

introducono nell'operazione di misura dei fattori soggettivi, cioè dipendenti dal modo di eseguire la misura da parte dell'operatore; questi fattori non si mantengono costanti al ripetersi dell'operazione di misura.

Questo causa una dispersione dei valori numerici che rappresentano il risultato delle misure.

Spesso si cerca di aumentare la precisione di uno strumento con operazioni di stima e di riporto, ma è proprio questa la causa degli gli errori.

Esempio: si vuole misurare la lunghezza di una trave rettilinea con un metro graduato in millimetri. Se la trave è lunga diversi metri, la variabilità dei risultati nasce almeno da due cause:

- a) siamo costretti a riportare più volte lo zero del metro;

b) se valutiamo la lunghezza al millimetro, dobbiamo stimare a quale tacca della graduazione corrisponde l'estremo della trave.

In ogni ripetizione del processo i **riporti** e la **stima** sono soggetti a fluttuazioni accidentali che generano perciò piccole variazioni nel valore finale stimato della lunghezza.

Notiamo che:

- tanto maggiore è il numero di riporti, tanto maggiori saranno le discordanze fra le ripetizioni della misura;
- se la trave fosse lunga meno di un metro, non occorrendo riporti, le misure differirebbero al più per 1mm.

Tali considerazioni ci impongono di abbandonare il concetto di valore che una grandezza come entità a se stante: si dovrà sempre esprimere il risultato di ogni operazione che misura associando al valore numerico la valutazione dell'incertezza con cui esso è stato ricavato.

A fianco al concetto di precisione si trova il concetto di accuratezza inteso come la concentrazione di misure ripetute. La misura ottimale dovrà ovviamente essere precisa ed accurata

Esistono, accanto alle fluttuazioni accidentali, anche le cosiddette cause sistematiche di errore, la cui natura emerge chiaramente considerando il modello usato per descrivere il fenomeno.

IL MODELLO

Ogni descrizione matematica di un fenomeno fisico, utilizzata per esprimere il valore che un a data grandezza in funzione di altre grandezze o parametri, deve ricorrere a semplificazioni.

Il **modello** deve essere perciò:

- a) il più semplice possibile, perché sia utilizzabile facilmente (non richieda la conoscenza o la misura di troppi parametri);
- b) complicato quanto necessario, in relazione alla approssimazione (incertezza) che si richiede ai valori predetti dal modello stesso.

Nel modello si distinguono una **componente funzionale** ed una **stocastica**, che sono strettamente connesse:

1) la **componente funzionale** descrive analiticamente la relazione fra la grandezza osservabile ed i parametri (fisici, geometrici) che sono ad essa collegati. La rilevanza, il numero ed il ruolo di questi parametri entro il modello deve essere valutato in relazione alla incertezza da ottenere nella stima della osservazione: in funzione di

tale valore, potranno assumere importanza o meno effetti di tipo sistematico che possono essere modellizzati.

2) la **componente stocastica** del modello è invece legata al complesso delle cause di variabilità del valore osservato che non si includono esplicitamente nel modello funzionale: essa tiene conto cioè della dispersione delle misure dovuta a cause, dette accidentali, che sfuggono ad una modellizzazione analitica o che si decide di non modellizzare analiticamente perché troppo complesse.

Esempio:

consideriamo la misura di una distanza piana L con una bindella centimetrata, lunga 50 m.

Il coefficiente di dilatazione termica della bindella è $b=10^{-5} \text{ C}^{-1}$ e la temperatura nell'ambiente di misura è di 20°C superiore a quella rispetto a cui la bindella è stata graduata. Si consideri che il coefficiente di allungamento del materiale costituente la bindella è $a = 5 \cdot 10^{-5} \text{ kg}^{-1}$ e che la tensione applicata in fase di misura è di 5 kg.

Le corrispondenti variazioni di lunghezza della bindella sono quindi pari a

$$\begin{aligned}\Delta L_1 &= b \cdot \Delta T = 50 \text{ m} \cdot 20 \cdot \text{C} \cdot 10^{-5} \cdot \text{C}^{-1} = 1 \text{ cm} \\ \Delta L_2 &= a \cdot F = 50 \text{ m} \cdot 5 \text{ kg} \cdot 5 \cdot 10^{-5} \cdot \text{kg}^{-1} = 1,25 \text{ cm}\end{aligned}$$

È evidente che se voglio misurare con incertezza vicina al cm devo tenere conto della deformabilità della bindella, cioè in sostanza devo correggere i valori misurati L_{oss} delle quantità L_1 e L_2 .

Il modello funzionale diventa perciò

$$L_{\text{oss}} = L \cdot (1 - b \cdot \Delta T - a \cdot F)$$

dove L_{oss} rappresenta la stima di L , cioè la valutazione della lunghezza ottenuta in base alle misure fatte.

Il modello funzionale include quindi due effetti sistematici (la dilatazione termica e l'allungamento dovuto alla forza applicata), considerati lineari, che richiedono di conoscere i valori di b ed a , oltre alla misura di T ed F .

Osservazioni:

1. In condizioni ambientali stabili e con forza applicata costante, se non correggo ottengo una lunghezza L_{oss} maggiore di L (lunghezza effettiva): commetto cioè un errore di stima della grandezza a cui sono interessato. La caratteristica di questo errore, che dipende da F e ΔT , è il fatto che posso prevederne l'entità, perché sono in grado di descriverlo analiticamente: **errore sistematico**
2. Supponiamo di voler misurare una lunghezza per cui sia necessario il riporto. Eseguiamo diverse serie di misure, variando in modo casuale da una serie all'altra le condizioni ambientali e le forze applicate, senza correggere le osservazioni. Ottengo allora una dispersione di risultati assai maggiore rispetto al caso precedente, in cui i

valori osservati sono divisi in tanti gruppetti, ciascuno corrispondente ad una serie di misura: se vario in modo accidentale le condizioni ambientali e la forza, queste cause assumono, se non corrette, un comportamento di tipo accidentale, poiché aumentano la dispersione delle osservazioni. Pertanto la distinzione fra comportamento sistematico ed accidentale, pur chiara concettualmente, all'atto pratico dipende dalle condizioni specifiche. Spesso, quando è presente un comportamento sistematico che tuttavia non si riesce a correggere (ad esempio perché non è noto il valore di un parametro da cui esso dipende), l'unica alternativa è proprio quella di rendere artificialmente accidentale il sistematismo, ripetendo più volte la misura.

Nelle operazioni topografiche gli strumenti impiegati, le condizioni operative e ambientali, le modalità di misura sono tali per cui l'incertezza di misura deve quasi sempre essere considerata. Si pone pertanto il problema di un corretto trattamento dei dati rilevati.

FENOMENI ALEATORI

Ci sono fenomeni il cui esito non è prevedibile a priori. Ad esempio:

- il risultato del lancio di un dado;
- la misura di una lunghezza;
- il peso di uno studente scelto a caso.

Per quanto incapaci di prevedere con esattezza il risultato del singolo evento, siamo però in grado evidenziare delle regolarità, di descrivere un comportamento "in media". Nel caso che più interessa il rilievo, ovvero la misura di grandezze, l'affermazione appena fatta implica l'assunzione che la variabilità (ovvero l'**incertezza**) di misura di tipo accidentale possa essere descritta a priori da un meccanismo di tipo probabilistico, cioè che le oscillazioni dei valori osservati siano rappresentabili come estrazioni da una **variabile casuale**.

Le discipline che studiano come descrivere e interpretare i fenomeni aleatori sono:

1. la **teoria della probabilità**: essenzialmente deduttiva, insegna a costruire le probabilità di eventi complessi a partire da un modello stocastico noto.
2. la **statistica**: di tipo induttivo, cerca di ricostruire un modello stocastico a partire da eventi già realizzati; essa si articola in:
 - **teoria della stima** (la ricerca della miglior strategia di interrogazione della realtà per estrarre informazioni sul fenomeno);
 - **inferenza**, cioè nella verifica di ipotesi sul modello interpretativo del fenomeno, verifica che, necessariamente, si effettua sulla base di dati "estratti" dal fenomeno.

Definizioni:

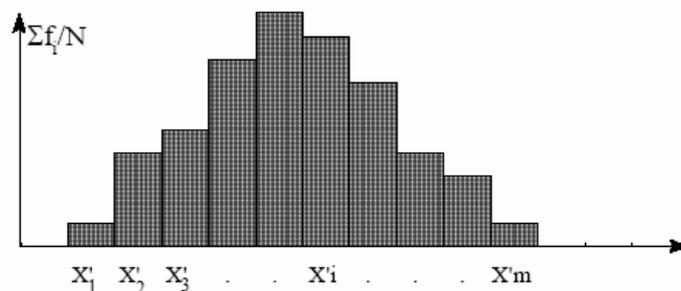
variabile aleatoria o stocastica: una entità, numerica o non, che può assumere uno solo tra tutti i valori di un insieme, ed a ciascuno dei quali è associata una probabilità.

variabile statistica: una entità che, nelle osservazioni effettuate sulle unità che costituiscono la popolazione, ha assunto più valori, frequentemente diversi tra loro.

Si definisce inoltre:

- a) *popolazione* l'insieme di N unità statistiche o individui che possiedono tutti una stessa caratteristica che si presenta in quantità differenti;
- b) *attributo* la caratteristica suddetta (oggetto della rilevazione);
- c) *valori argomentali* i differenti valori dell'attributo che possono presentarsi negli individui della popolazione;
- d) *frequenza assoluta* il numero degli individui che hanno lo stesso valore argomentale;
- e) *frequenza relativa* il rapporto tra la frequenza assoluta ed il numero totale degli individui della popolazione.

Una variabile statistica può essere rappresentata graficamente mediante l'istogramma.

La ripetizione delle misure o ridondanza

E' necessario sottolineare che condizione necessaria per applicare un trattamento statistico alle misure è la ripetizione delle osservazioni (ridondanza).

Analiticamente la ridondanza è data da:

$$r = n - m$$

con n = numero delle osservazioni e

m = numero delle incognite

CLASSIFICAZIONE DEGLI ERRORI

Come abbiamo visto se misuriamo una grandezza più volte, otterremo una serie di misure, che, causa gli errori sempre presenti nell'atto del misurare, differiranno fra loro.

1- **errori grossolani**: quando in una serie di misure si ha un valore che si discosta nettamente. Sono dovuti essenzialmente alla disattenzione o alla poca perizia dell'osservatore. Si mettono in evidenza e si eliminano eseguendo un numero esuberante di misure per poter in seguito eseguire delle verifiche (esempio; trascrizione errata di una misura,...);

2- **errori sistematici o biases**: si manifestano regolarmente al ripetersi delle misurazioni e sono da imputare a varie cause, quali difetti strumentali. Per correggerli

si ricorre a **leggi deterministiche** (errori di verticalità, errori nel considerare lo zero della cordella metrica,...)

3- **errori accidentali o random errors**: si presentano con ammontare diverso ad ogni misura, spesso non prevedibili. Non sono eliminabili ma riducibili aumentando la precisione dello strumento, oppure ripetendo l'operazione. Per correggerli si ricorre a **leggi statistiche** e costituiscono il campo di applicazione della **teoria degli errori**, la quale ha come duplice scopo quello di:

- a) trarre una serie di osservazioni di una grandezza un valore da assumere che, tra tutti quelli deducibili, dia il massimo affidamento, ossia abbia la massima probabilità di coincidere con un valore vero;
- b) stabilire un criterio globale per giudicare la precisione delle osservazioni eseguite e il valore assunto dalla grandezza osservata, avendo constatato che gli errori accidentali non si eliminano, ma si **compensano**.

PARAMETRI STATISTICI

Considereremo le nostre serie di misure affette da soli errori di tipo accidentale.

Avere una serie di dati iniziali comporta la necessità di trovare certe caratteristiche numeriche che riassumono fedelmente le informazioni contenute nei dati e che si prestino bene ai calcoli.

Questi valori, destinati a rappresentare i dati iniziali di una serie di misure, devono permettere calcoli ulteriori e rappresentare i dati iniziali; essi vengono appunto detti **parametri statistici** della serie in oggetto.

Un parametro statistico è tanto più efficace quanto meglio riassume il contenuto informativo dei dati iniziali con la minor perdita di informazioni e quanto meglio si presta a calcoli ulteriori e test.

I parametri statistici più efficaci sono:

- la **media aritmetica** dei dati;
- la **varianza** o la **deviazione standard**.

Questi due parametri sono quelli che contengono la maggior quantità di informazioni utili.

La **MEDIA** è il valore centrale attorno a cui oscillano i valori trovati.

$$\mu = \frac{1}{N} \sum_i X_i \cdot f_i$$

Essa racchiude solo una parte dell'informazione sui dati perché le manca una caratteristica essenziale, quella cioè di non dirci come i valori oscillano intorno a lei. Viene definita un indice di posizione perché indica la posizione del valore centrale sulla scala dei valori.

Esistono altri indici di posizione come:

la **MODA**, che rappresenta il valore più frequente della distribuzione
 la **MEDIANA** che rappresenta il valore centrale della distribuzione, come risulta dopo il riordino della popolazione.

La **VARIANZA** misura la dispersione dei valori attorno alla media, e deriva da considerazioni matematiche. La varianza, che si indica generalmente con la σ^2 , la definiamo matematicamente come la somma dei quadrati degli scarti divisi per il numero dei gradi di libertà.

$$\sigma^2 = \frac{1}{N} \sum_i (x_i - \mu)^2 \cdot f_i$$

A differenza della media, la varianza viene definita un indice di dispersione, perché indica come le osservazioni sono concentrate attorno ad un valore.

Un altro parametro intuitivo è lo **SCARTO** (o deviazione standard), ovvero la differenza fra un generico valore della serie e il valore medio.

La **DEVIAZIONE STANDARD** o **SQM** altri non è che la radice σ della varianza.

Riassumendo:

Media	$\mu = \frac{1}{N} \sum_i X_i \cdot f_i$
Varianza	$\sigma^2 = \frac{1}{N} \sum_i X_i^2 \cdot f_i$
Deviazione standard o sqm	$\sigma = \sqrt{\sigma^2}$
Errore standard della media o <i>eqm</i> della media	$\sigma_m = \frac{\sigma}{\sqrt{N}}$

dove f = frequenza relativa

N = popolazione

X = osservazione

x = scarto

Esempio:

calcolo del valore dell'altezza di un vano.

Nel nostro caso è:

Popolazione: 7 misure

Attributo: altezza del vano

Valori argomentali: 2.715; 2.70. 2.69. 2.705; 2.71; 2.685; 2.70

Osservazioni X_i [m]		Scarto x_i ($X_i - \mu$) [m]		Scarto al quadrato x_i^2	
	2,725		0,022		0,00049
	2,705		0,002		0,00000
	2,690		-0,013		0,00017
	2,715		0,012		0,00015
	2,700		0,003		0,00001
	2,685		-0,018		0,00032
	2,700		-0,003		0,00001
sommatoria Σ :	18,920	sommatoria Σx_i :	0,000	sommatoria Σx_i^2 :	0,00114
media μ [m]=	2,703			varianza σ^2 [m ²] =	0,00019
moda [m]=	2,700			sqm σ [m]=	$\pm 0,0106$
mediana [m]=	2,700				

I gradi di libertà

Si definisce Serie di misure **indipendenti**: una serie in cui nessun valore può essere dedotto dalla conoscenza degli altri valori. Sono cioè indipendenti fra loro. Se le misure sono n e indipendenti, i gradi di libertà sono parimenti n.

Si considerino gli scarti x_i : in una serie l'ultimo è determinato e può essere calcolato a partire dai primi valori perché la somma algebrica degli scarti deve essere uguale a zero. Perciò, nella serie di n scarti n-1 valori sono indipendenti fra loro, ovvero i gradi di libertà della serie scarti sono n-1.

Passando dalla serie misure alla serie scarti si perde sempre un grado di libertà. ·

⇒ I gradi di libertà esprimono il numero di dati effettivamente disponibili per valutare la quantità di informazioni contenuta in essi. ·

⇒ Non è il numero totale dei dati che conta, ma il numero dei dati indipendenti. quando un dato non è indipendente le sue informazioni potenziali sono già presenti negli altri dati.

Una serie di misure come campione di una propagazione

Una serie di misure altri non è che un campione preso dalla popolazione delle misure possibili di una grandezza, che sono infinite. Fino a che punto il campione esprime le caratteristiche della popolazione originaria?

L'esperienza ha dimostrato che la maggior parte delle misurazioni possono considerarsi estratte da popolazioni di origine distribuite normalmente.

In una distribuzione normale:

- ⇒ uno dei valori apparirà con frequenza massima,
- ⇒ i valori più bassi o più alti di questo compariranno con una frequenza tanto minore quanto si allontanano dal valore più frequente.

Risulta lecito porsi il quesito di trovare un tipo di funzione $f(x)$ che interpoli bene gli istogrammi (andamento variabile causale normale) e che possa essere considerata come la funzione della distribuzione delle probabilità della variabile causale di media μ e varianza σ associabile ad una qualsiasi popolazione di misure possibili.

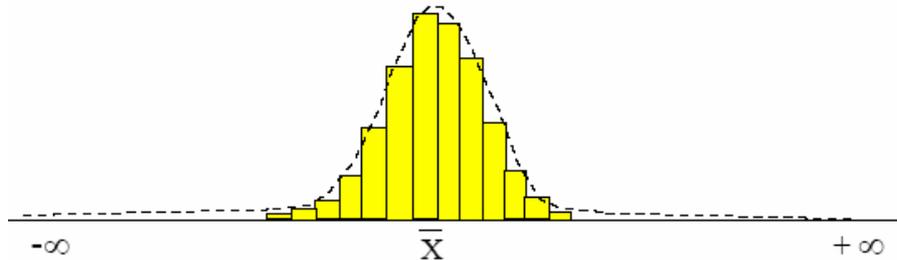
La funzione $f(x)$ che risulta più idonea a questo scopo, è la curva di Gauss, che ha la seguente espressione:

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

dove μ e σ sono rispettivamente media e varianza della variabile causale.

Possiamo pertanto concludere che:

- dati una certa quantità di grandezza da misurare, lo strumento di misura e l'ambiente in cui si opera, si genera una popolazione di misure possibili;
- questa popolazione di misure è rappresentabile con una variabile casuale continua, definita tra $-\infty$ e $+\infty$, che ha una media μ , uno scarto quadratico medio σ ed una distribuzione di probabilità definita dalla espressione della curva di Gauss.



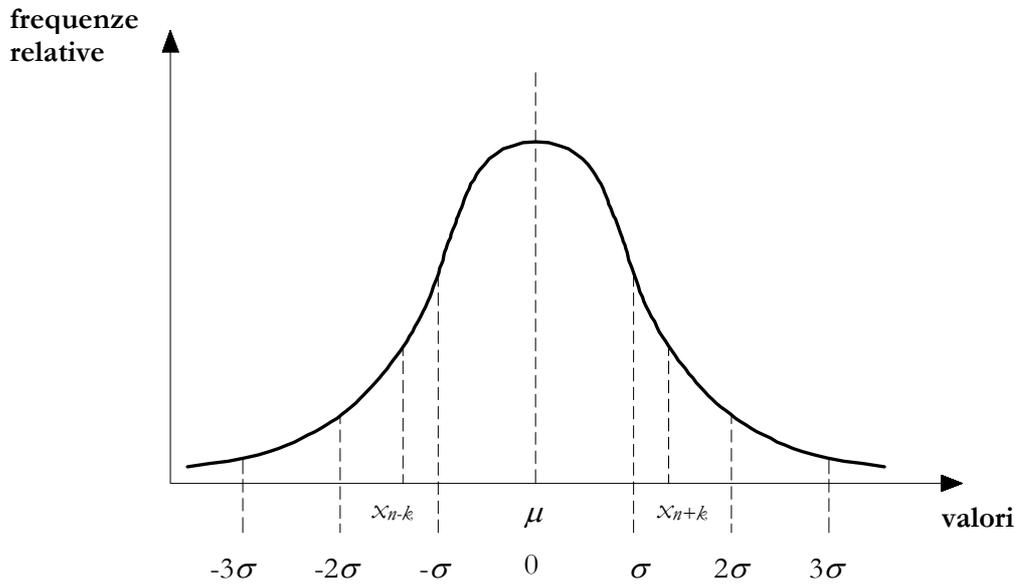


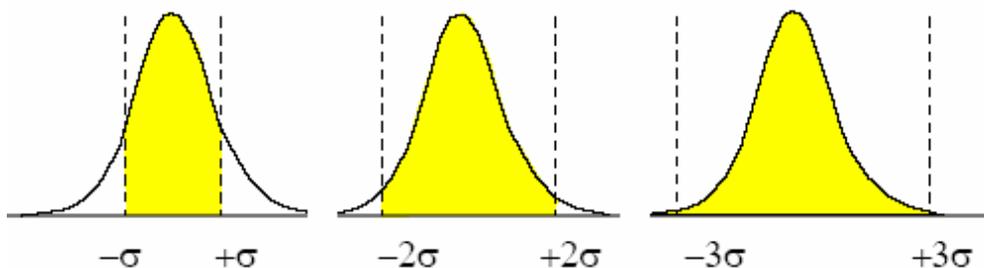
Fig 1: curva gaussiana e distribuzione degli errori

La curva normale, detta curva o distribuzione di GAUSS ha una forma a campana ed è simmetrica rispetto al valore di massima frequenza. E' compiutamente nota quando si conoscono l'ascissa della sommità della curva, che è il valore medio, e la distanza da questo dei punti di flesso della curva, simmetrici a destra e a sinistra della media, che altri non sono che la deviazione standard.

La figura mostra la distribuzione normale di una popolazione con media μ e deviazione standard $\pm \sigma$. I valori che si allontanano dalla media sono più rari di quelli ad essa vicini.

L'espressione matematica della curva, che è asintotica, ci dice che tutti gli individui della popolazione stanno sotto la curva tra $-\infty$ e $+\infty$ + la probabilità che un individuo preso a caso fra la popolazione presenti un valore compreso entro un intervallo assegnato è data dal calcolo dell'area sottesa dalla curva in quel intervallo. Questi valori sono contenuti in apposite tavole.

Per esempio il 68,26% della popolazione si trova nell'intervallo $\mu \pm \sigma$, il 94,44% fra $\mu \pm 2\sigma$, il 99,73% fra $\mu \pm 3\sigma$, il 100% fra $\mu \pm \infty$.



La stima di μ e di σ

Si può dedurre che è possibile stimare i valori più probabili di μ e σ .

- La miglior stima della media μ della popolazione è la media aritmetica \bar{X}
- La miglior stima della deviazione standard σ della popolazione è la deviazione standard S del campione.

L'errore standard della media

Negli esercizi precedenti si sono rappresentati i parametri statistici \bar{X} e S del campione estratto dalla popolazione delle misure possibili. Supponiamo ora di estrarre dalla popolazione un altro campione di sei misure e calcoliamo la media \bar{X}_1 (pure essa una stima di μ): il suo valore sarà leggermente diverso da \bar{X} .

Procediamo analogamente ottenendo un altro campione di sei individui e quindi una media \bar{X}_2 e così via. Se ripetessimo infinite volte questa operazione otterremmo una popolazione di campioni ciascuno con il suo valore della media. Esiste un teorema fondamentale che dice che : "Se una popolazione è distribuita normalmente con media μ e deviazione standard σ , le medie di un numero infinito di campioni, ciascuno composto da n individui estratti a caso dalla popolazione, si distribuiscono secondo la curva normale la cui media è μ e la cui deviazione standard è $\sigma_m = \sigma/\sqrt{n}$.

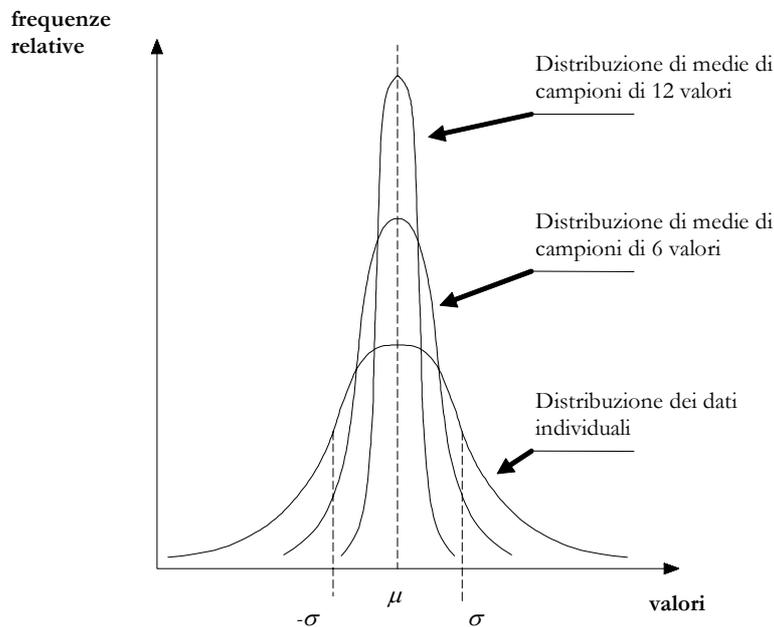


Fig 2: Curva di Gauss: individui e medie dei campioni

Le medie calcolate a partire da un campione oscillino molto meno attorno alla media

di quanto non facciano gli individui che costituiscono il campione. Quanto è maggiore la dimensione del campione ovvero il numero n di individui che lo compongono, tanto minore è la dispersione delle medie attorno al valore vero. Al limite, quando $\mu \pm \infty$, la deviazione standard tende a zero e quindi la media stimata tende alla media vera ovvero alla misura "vera".

SCARTO QUADRATICO MEDIO E TOLLERANZA

Data quindi una serie di osservazioni, il calcolo dello scarto quadratico medio s.q.m.,

oltre a fornire il grado di precisione, tramite l'equazione $\sigma = \pm \sqrt{\frac{\sum x_i^2}{n-1}}$ con cui si sono effettuate le misure, stabilisce anche i limiti, dati dall'intervallo $\mu - 3\sigma < X_1 < \mu + 3\sigma$, oltre i quali gli errori non potranno più essere considerati come accidentali e pertanto trattati come tali.

Per questo motivo il valore:

$$t=3\sigma$$

è chiamato **tolleranza** e tutte le osservazioni, i cui scarti superano tale valore, dovranno essere scartate dalle misurazioni in quanto probabilmente affette da errori materiali (grossolani).

Pertanto fissare la tolleranza in un rilievo significa fissare la massima quantità in cui una misura può discostarsi dal valore della media (valore più probabile) in pratica significa fissare l'ampiezza dell'intervallo di dispersione delle misure.

Se, ad esempio, viene richiesto di determinare la quota di un punto con la tolleranza di ± 2 cm, questo significa che si dovrà operare con una metodologia tale per cui la popolazione di misure possibili sia tutta contenuta entro un intervallo che va da -2 cm a +2 cm nell'intorno del valore μ , che possiamo chiamare quota "giusto" del punto in questione.

Ma, per ottenere questo risultato, le operazioni dovranno essere ottimizzate per ottenere un s.q.m. pari a $\pm \frac{2}{3}$ cm.

Generalmente di un metodo di misura o di uno strumento si dà l's.q.m. a priori. Si dice ad esempio: con questo teodolite si possono misurare gli angoli con s.q.m. $\pm 2''$ di errore, intendendo con questo che si può sbagliare nella determinazione dell'angolo anche di $6''$. Imporre l's.q.m. significa imporre anche le condizioni di accidentalità degli errori e della loro distribuzione gaussiana. Viceversa se si impone la tolleranza si deve accettare un lavoro anche se si constata che gli errori in esso presenti, pure essendo tutti inferiori alle tolleranze, sono tutti dello stesso segno e, come valore assoluto, tutti poco al di sotto della tolleranza.

PROPAGAZIONE DELL'ERRORE

Abbiamo visto come la media e la varianza di una variabile casuale monodimensionale siano in grado di rappresentarne il baricentro e la dispersione. In topografia però raramente si misura direttamente la quantità che si vuole determinare: ad esempio si misurano angoli azimutali, zenitali, distanze e dislivelli per determinare coordinate ecc.(osservazioni indirette).

Occorre allora poter ricavare la media e la varianza della variabile casuale funzione di altre variabili casuali, ovvero derivare le caratteristiche di aleatorietà della grandezza misurata indirettamente, una volta noti i parametri caratteristici delle quantità misurate direttamente.

Nel caso di **osservazioni lineari indipendenti** l'una dall'altra, la legge che regola la propagazione dell'errore può essere scritta in questo modo:

$$\sigma_f^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 + \dots$$

in cui σ_x^2 e σ_y^2 sono le varianze delle diverse misure, mentre a e b sono i coefficienti.

Esempio:

calcolo dell'incertezza di una misura

Sommando il segmento AB di lunghezza 125 con incertezza pari a 3 cm e il segmento BC di lunghezza 154 e incertezza pari a 4 cm, quale sarà la lunghezza e l'incertezza del segmento totale AC?

La lunghezza sarà data da: $125 + 154 = 279 \text{ cm}$

Mentre l'incertezza sarà: $3^2 + 4^2 = 25 = \pm 5 \text{ cm}$

per cui il segmento AC = 279 cm +/- 5 cm

IL PRINCIPIO DEI MINIMI QUADRATI

Supponiamo di avere una grandezza η della quale si facciano n osservazioni l indipendenti; siano v_1, v_2, \dots gli errori di osservazione.

$$\begin{aligned} v_1 &= l_1 - \eta \\ v_2 &= l_2 - \eta \\ &\dots \\ v_n &= l_n - \eta \end{aligned}$$

Il problema è quello di determinare il valore più probabile di η sulla base delle n osservazioni l , il che equivale a trovare i più probabili valori per i v_i . Questi ultimi, nelle ipotesi fatte, sono n variabili casuali, ciascuna delle quali segue la distribuzione normale, con densità di probabilità data dalla

$$f(v_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{v_i}{\sigma_i}\right)^2}$$

Siccome le variabili sono fra loro indipendenti per il teorema delle probabilità composte, la probabilità che avvengano n eventi contemporaneamente è data dal prodotto delle probabilità dei singoli eventi, si ha che la funzione densità di probabilità congiunta è data dal prodotto delle varie funzioni componenti.

$$\begin{aligned} f(v_1, v_2, \dots, v_n) &= \frac{1}{\sigma_1 \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{v_1}{\sigma_1}\right)^2} \cdot \frac{1}{\sigma_2 \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{v_2}{\sigma_2}\right)^2} \cdot \dots \cdot \frac{1}{\sigma_n \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{v_n}{\sigma_n}\right)^2} = \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma_1} \cdot \frac{1}{\sigma_2} \cdot \dots \cdot \frac{1}{\sigma_n}\right) \cdot e^{-\frac{1}{2} \sum_1^n \left(\frac{v_i}{\sigma_i}\right)^2} \end{aligned}$$

La funzione $f(v_1, v_2, \dots, v_n)$ densità di probabilità raggiunge il massimo del suo valore quando l'esponente raggiunge il minimo. Per cui si può scrivere:

$$\sum_1^n i \left(\frac{v_i}{\sigma_i}\right)^2 = \min$$

Se introduciamo una costante arbitraria σ_o , non viene modificato il risultato della condizione di minimo, in quanto σ_o rappresenta numericamente un fattore di scala.

$$\sum_1^n i \frac{\sigma_o^2 v_i^2}{\sigma_i^2} = \min$$

Detto $P_i = \frac{\sigma_o^2}{\sigma_i^2}$, peso dell'osservazione, la condizione di minimo diventa:

$$\sum_1^n i P_i v_i^2 = v^t P v = \min$$

Una osservazione che abbia varianza $\sigma_i^2 = \sigma_o^2$ ha peso 1 e perciò è chiamata *varianza dell'unità di peso*.

Usando altri termini, si può affermare che data una serie di osservazioni indipendenti, i valori più probabili sono quelli per cui la sommatoria dei loro scarti al quadrato sia minima.

Questo tipo di compensazione riguarda sistemi di equazioni lineari o linearizzate che hanno un numero di equazioni maggiore del numero di variabili. I sistemi sovradeterminati (più equazioni che incognite) non hanno una soluzione unica e quindi si cerca la soluzione migliore (di minima varianza). A causa degli errori di misura presenti, errori che riterremo di tipo accidentale ovvero escludendo errori sistematici o grossolani, il legame tra osservazioni e incognite avrà un residuo v . In poche parole, se risolvessimo nel sistema sovradeterminato un numero di equazioni uguale al numero delle incognite x avremmo un sistema esattamente determinato. Sostituendo però le incognite x nelle rimanenti equazioni del sistema sovradeterminato queste non sono identicamente nulle, ma ammettono dei residui v dovuti agli errori di misura.

Conviene pertanto determinare le incognite x in modo tale che i residui siano ripartiti in ragione proporzionale all'incertezza delle osservazioni, così che sia minima la sommatoria dei loro scarti quadrati.

Il sistema sovradeterminato è del tipo $f(x, \alpha) = v$, dove ogni equazione alle misure è del tipo $f_i(x, \alpha) = v_i$.

Vi sono due problemi da risolvere.

Il primo è che non tutte le equazioni del sistema "contano allo stesso modo", occorre cioè *pesarle*.

Il secondo è che molto spesso le equazioni si presentano in forma non lineare e perciò occorre linearizzarle.

La soluzione al secondo problema è semplice in quanto se supponiamo che le quantità misurate α_i non siano eccessivamente disperse, ovvero siano affette da errori solo accidentali e piccoli, anche le x_k non differiranno eccessivamente tra di loro.

Sarà perciò possibile trovare dei valori approssimati delle incognite x e operare un cambiamento di variabili del tipo $x = x_o + \delta x$ sviluppabile in serie di Taylor arrestata al primo ordine.

Esplicitando una equazione $f_i(x, \alpha) = v_i$ generica del sistema si ha:

$$f_i(x_o, \alpha) + \left[\frac{\partial f_i}{\partial x_1} \right]_o \delta x_1 + \left[\frac{\partial f_i}{\partial x_2} \right]_o \delta x_2 + \dots + \left[\frac{\partial f_i}{\partial x_m} \right]_o \delta x_m = v_i$$

dove il primo termine è una quantità nota l_i e i coefficienti

$$\left[\frac{\partial f_i}{\partial x_k} \right]_0$$

delle nuove incognite δx_k costituiscono una matrice A .

Il nuovo sistema lineare (o linearizzato) assume la forma $A\delta x + l = v$ in m incognite δx e n incognite v .

Il primo problema, quello dei pesi delle equazioni, può essere ricondotto alla determinazione della varianza dell'osservazione. Anche senza doverla stimare empiricamente attraverso l'analisi della distribuzione di ogni singola osservazione, si può fare riferimento, nell'ambito delle misure, alla varianza dello strumento di misura. Il peso da assegnare a ciascuna equazione è perciò inversamente proporzionale alla varianza dell'osservazione.

La matrice dei pesi P è una matrice diagonale del tipo:

$$P = \begin{vmatrix} P_1 & 0 & \dots & 0 \\ 0 & P_2 & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & P_n \end{vmatrix}$$

Il sistema sottodeterminato $A\delta x + l = v$, associato alla condizione di minimo pesato della somma degli scarti al quadrato, dà luogo ad un sistema esattamente determinato detto *sistema normale*

$$\begin{cases} A\delta x + l = v \\ \sum P v^2 = \text{minimo} \end{cases} \Rightarrow \text{sistema normale}$$

la cui soluzione è data da:

$$\delta x = \begin{vmatrix} \delta x_1 \\ \delta x_2 \\ \dots \\ \delta x_m \end{vmatrix} = -(A^t P A)^{-1} A^t P l$$

La stima delle varianze delle incognite δx è data dai termini diagonali h_{ii} della matrice quadrata $(A^t P A)^{-1}$ moltiplicati per un termine σ_0^2 detto varianza dell'unità di peso.

Questo è formato dalla sommatoria degli scarti pesati al quadrato divisi per la ridondanza del sistema, ovvero

$$\sigma_0^2 = \frac{\sum P v^2}{n - m} = \frac{v^t P v}{n - m}$$

LA PRECISIONE NEGLI ELABORATI DI RILIEVO: ERRORE DI GRAFICISMO E SCALA NOMINALE

A questo punto si può affermare che il primo obiettivo delle operazioni di rilievo non è la determinazione del valore vero della misura, ma stabilire statisticamente l'errore che potremmo incontrare nelle misurazioni e quindi prefigurare e fissare l'intervallo all'interno del quale considerare attendibile il risultato.

In ogni operazione di rilievo andrà così valutata l'incertezza che si potrà riscontrare nelle misure, ed a lavoro effettuato andrà effettivamente verificato che non si sia superato quel valore.

Fino ad ora si è osservato come l'incertezza dipenda da molti parametri, tra cui le condizioni in cui si opera, le procedure, le tecniche, gli strumenti, l'operatore...

Sperimentalmente è stato comunque verificato che è possibile utilizzare come parametri di incertezza quelli derivanti dall'*errore di graficismo*, in cui le incertezze sono relazionate alla scala di rappresentazione.

Per convenzione l'errore di graficismo è l'errore insito in una linea dello spessore di 0.2mm (lo spessore più piccolo comunque riproducibile al tratto), per cui ad ogni scala questo fattore assume un valore specifico.

Si calcola moltiplicando 0.2 mm per il denominatore della scala

$$\text{e.g.} = 0.2 \text{ mm} \times \text{denominatore Scala}$$

Se ad esempio considero un rilievo in scala 1:100, una linea di 0,2 mm rappresenta lo spessore di 2 cm. Se misuro un tavolo su quel disegno, leggo sul decimetro il valore di 1.8 cm, che riportato in scala corrisponde a 1,8 metri. Questo valore è misurato considerando le due linee di bordo, per cui non sapendo cosa succede nello spessore al reale delle linee, per esattezza dovrei dire che il mio tavolo è lungo 1.8 ± 0.02 metri.

Un'ulteriore considerazione riguarda la relazione tra l'incertezza e l'attuale produzione degli elaborati di rilievo. L'uso dello strumento informatico è ormai completamente assodato e diffuso, per cui è necessario anche relazionarsi a questo strumento. Trattandosi di disegno al CAD, generalmente si ritiene che non esistano limiti di scala e che si possa disegnare – e quindi rilevare – a qualsiasi scala, e solo in fase di stampa decidere il fattore di riproduzione. Non esistono più neppure i limiti dati dallo spessore della matita che una volta impedivano il tracciamento di due linee parallele oppure rendevano incerto il punto d'incontro di due archi di cerchio. Per questi motivi sembra ormai improprio parlare di errore di graficismo. Questo non è del tutto vero perché anche se sono stati eliminati gli errori di tracciamento (due archi di cerchio di incontrano in un punto univoco), il disegno al calcolatore non interviene nella precisione della misurazione che si effettua sul foglio stampato. Grazie a quest'ultima considerazione si considera valido il concetto di errore di graficismo e gli si affianca anche il concetto di **SCALA NOMINALE**, intesa come la scala di riduzione (o di

ingrandimento) per cui è corretto stampare un disegno. Quando si dice che un rilievo restituito al CAD ha scala nominale 1:N si intende che ha contenuto metrico e semantico di un corrispondente disegno su carta di pari scala; per cui l'incertezza delle misure e il numero dei segni risulta appropriato.

Viene utilizzato il termine nominale per contraddistinguerla dalle normali scala di rappresentazione. Infatti avendo un rilievo realizzato al CAD, è sempre possibile stamparlo con diversi fattori di scala: in alcuni casi il disegno stampato sarà troppo ricco di segni (e quindi le linee si impasteranno tra loro), mentre in altri casi il disegno sarà troppo povero.

La scala nominale non fornisce solamente il valore dell'incertezza del disegno, ma da anche delle indicazioni all'operatore ancor prima di iniziare la fase di misurazione. Infatti è inutile rilevare i dettagli che sono contenuti nell'incertezza del disegno, perché risulteranno "interni" allo spessore della linea. Così ad esempio dovendo rilevare un dettaglio, o un ordine architettonico, alla scala 1:100, non si dovranno rilevare gli elementi di dimensione minore al centimetro.

Se si realizza un rilievo a scala nominale 1:100 (e quindi si rilevano tutti gli elementi maggiori o uguali a 2 cm) e lo si stampa ad una scala minore 1:200 o 1:500, le linee saranno impastate e non sarà possibile riconoscerle con esattezza. Se invece si stampa lo stesso disegno in scala 1:50 o 1:20, il disegno sarà povero di linee.

Scala NOMINALE	INCERTEZZA
1:1000	20 cm
1:500	10 cm
1:200	4 cm
1:100	2 cm
1:50	1 cm
1:20	0,4 cm (4 mm)
1:10	0,2 cm (2 mm)
1:5	0,1 cm (1 mm)

Tab. 1: rapporto tra scala nominale e incertezza