### CAPITOLO III – DATI E METADATI

Ogni informazione scientifica si dovrebbe costruire a stadi, definendo dapprima il *metadato* che la qualifica e poi eseguendo la misura che la quantifica (dato): il metadato essendo formato dall'insieme delle notizie (unità di misura, strumento usato per ottenerlo, modalità di campionamento, luogo e data di misura, ecc.) che servono a stabilire quale sia il significato da attribuire al numero che costituisce il dato. Purtroppo, non sempre si procede in questo modo e, così, spesso il dato finisce per perdere scientificità perché diventa non verificabile. Inoltre, si deve fare attenzione al termine perché, semplificando, spesso si usa la parola "dato" per indicare l'informazione completa (dato e relativo metadato) o, peggio, per indicare un'informazione pseudoscientifica, cioè priva delle informazioni necessarie per verificarla. In ogni caso, sospetti e comunque poco affidabili appaiono i dati conservati in un cassetto con l'obiettivo (dichiarato o meno) di tirarli fuori all'occorrenza. Ovvero, caratteristica accessoria, ma non marginale per la valenza scientifica dei dati ambientali è la loro pubblicità.

### 3.1. – Strutture di dati

Con riferimento all'ambiente, le informazioni si possono classificare in relazione alla loro struttura come:

❖ DATI per indicare informazioni puntuali, locali e monotematiche (ad esempio, la temperatura istantanea dell'acqua rilevata in un punto di un fiume). Questi dati costituiscono l'informazione ambientale elementare che risulta composta dal valore numerico o qualitativo di un indice e dal relativo metadato. Talvolta si classificano come *dati* anche informazioni composite come nel caso dell'archivio SYNOP il cui tracciato record è riportato in fig.3.1 che mostra l'articolazione

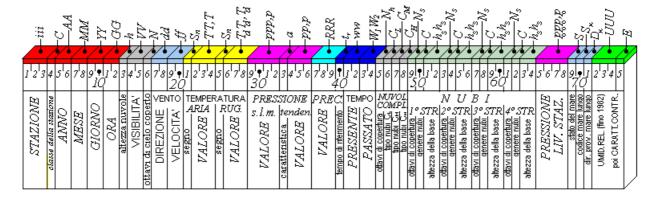


Fig. 3.1. - Tracciato record dell'archivio meteorologico SYNOP.

del concetto di tempo meteorologico secondo l'Organizzazione Meteorologica Mondiale.

- ❖ SEQUENZE di dati, spesso indicate col nome di serie temporali, che sono raccolte di dati relativi allo stesso argomento ma riferiti a tempi successivi;
- ❖ SCENE che riferiscono la stessa informazione raccolta nello stesso istante in punti diversi del territorio. Si tratta generalmente di immagini fotografiche o numeriche.
- ❖ EPISODI che sono costituiti da sequenze di scene relative allo stesso territorio.

La ricchezza e la varietà delle informazioni necessarie per studiare l'ambiente ed i processi che lo modificano è tale da renderne improponibile la gestione manuale. Per questo, ma solo da poco, si stanno sviluppando le archiviazioni elettroniche che facilitano sia la consultazione automatica con il supporto di efficienti "motori di ricerca", sia la condivisione dei dati attraverso le reti informatiche. Di fatto, problemi politici di *privacy* e di proprietà stanno ritardando la completa liberalizzazione delle banche di dati ambientali in nome di un dibattito sulle modalità di accesso in corso fra quanti auspicano la totale gratuità e quanti ne sostengono la completa riservatezza. Fortunatamente sull'argomento la legislazione nazionale e quella internazionale si stanno muovendo nella direzione della circolazione libera di tutti i dati ambientali di produzione pubblica (v. ad es. D.L. 24 febbraio 1997 n. 39 e la Direttiva CEE 1990/313 che ha recepito).

#### 3.2. – Natura discreta dei dati

Benché sia possibile immaginare e costruire sensori capaci di misurare in continuo molte grandezze, di fatto i dati su cui poggia la conoscenza scientifica sono sempre di natura discreta. I numeri irrazionali (quelli che derivano da rapporti fra numeri primi o da incommensurabilità di figure come per la quadratura del cerchio) hanno valenza teorica, ma inconsistenza pratica perché si

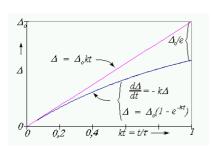


Fig. 3.2. – Smorzamento critico di un misuratore.

continuerà a dire che  $\pi$  vale 3,14 e che 1/3 vale 0,3. D'altra parte, gli strumenti, anche i più sofisticati, quando vengono usati per misurare grandezze variabili nel tempo pongono dei problemi di precisione che sono di fatto problemi di significatività. Si immagini un bulbo termometrico immerso nell'atmosfera. La capacità termica dell'elemento sensibile conferisce un'inerzia alla risposta dello strumento. Il *lag* temporale fra una misura e la successiva (l'intervallo di tempo che occorre al bulbo inizialmente ad una temperatura per mettersi in equilibrio con l'atmosfera ad una temperatura diversa) definisce la *«costante di tempo»* dello strumento (per

maggiori dettagli v. Appendice). Ovvero, uno strumento risponde alle variazioni della grandezza che sta misurando seguendo una legge esponenziale perché è stato costruito con uno smorzamento prossimo a quello critico (quello, appunto che raggiunge asintoticamente il valore da misurare impiegando il minor tempo). In fig.3.2 è illustrato il funzionamento tipico degli strumenti di misura insieme con i significati di *lag temporale* o *tempo di salita* dello strumento ( $\tau$ ) e di costante di tempo, k (=  $1/\tau$ ). I due parametri sono caratteristiche costruttive dello strumento e ne misurano l'incapacità a rilevare variazioni più veloci di k (lo strumento funziona da filtro passa-basso) e quindi non stabilisce l'esatto valore della variabile da misurare, ma lo approssima eliminando le più alte frequenze del segnale a rischio anche di alterarne il significato.

## 3.3. – Misurazioni e campionamenti

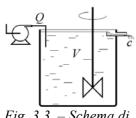


Fig. 3.3. – Schema di un miscelatore.

Con riferimento ad un termometro si è detto come, in generale, un misuratore possa determinare con continuità tutti i valori che il parametro da misurare è in grado di assumere, ma si è anche asserito che in condizioni dinamiche o di non equilibrio (disuniformità spaziale) la misurazione crea alcuni problemi. Si voglia, ad esempio, determinare la concentrazione di una sostanza che esce da un miscelatore di volume V alimentato da una pompa a portata costante Q, contaminato istantaneamente al tempo t=0 con una quantità M di tracciante (v. fig.3.3). La miscelazione si immagina completa (in ogni istante la

concentrazione del tracciante è uniforme in tutto V). Un semplice bilancio stabilisce che in prima approssimazione (la relazione che si ottiene è tanto più vera quanto più piccolo è l'intervallo di tempo  $\Delta t$  considerato) la massa di tracciante  $M(t+\Delta t)$  presente all'istante  $t+\Delta t$  è pari alla massa M(t)

presente all'istante t diminuita della massa uscita nell'intervallo di tempo  $\Delta t$  che è pari a  $Qc(t)\Delta t$ , cioè:

$$M(t + \Delta t) = M(t) - Qc(t)\Delta t$$

dove c(t) indica la concentrazione del tracciante all'istante t. Per la precisione, il bilancio è esatto al limite per intervalli di tempo molto piccoli (tendenti a zero:  $\Delta t \rightarrow 0$ ), bilancio che allora si scrive in termini differenziali nella forma:

$$dM/dt = -Qc$$
, ovvero  $dc/dt = -Qc/V = -c/\tau$ 

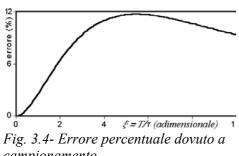
dove  $\tau = V/Q$  è il tempo di residenza o tempo di permanenza dell'acqua nel miscelatore.

In termini integrali si ottiene:

$$c(t) = c_o \exp(-t/\tau)$$

dove  $c_o = M/V$  è la concentrazione iniziale del tracciante.

Questa equazione si presta a considerazioni semplici, ma di notevole efficacia per valutare gli effetti del campionamento sulle misure, effetti che in campo ambientale sono generalmente



campionamento.

prevalenti rispetto a quelli dovuti alla precisione degli strumenti usati. I campionamenti possono "istantanei" (ad esempio con procedimenti fotometrici si raggiungono tempi di esposizione inferiori ai milionesimi di secondo) o prolungati nel tempo (esistono campionatori passivi per esposizioni anche di anni). In ogni caso il campionamento è caratterizzato da un tempo di esposizione T (i campionatori sono anche caratterizzati dalle dimensioni spaziali del sensore, ma tenerne conto in questa sede complicherebbe solo i problemi senza niente aggiungere al ragionamento). La concentrazione di

tracciante di un campione prelevato all'istante t nel miscelatore di fig.3.3 (o alla sua uscita: l'ipotesi di miscelazione completa assicura l'invarianza del campione rispetto alla posizione di prelievo) ha il valore  $\underline{c}_t$  della media temporale di c(t) nell'intervallo fra  $t \in t + T$ , cioè:

$$\underline{c}_t = \frac{1}{T} \int_t^{t+T} c(t) dt = \frac{\tau}{T} c(t) \left\{ 1 - e^{-\frac{T}{\tau}} \right\}$$

e la differenza  $\Delta c = \underline{c_t} - c(t + T/2)$  fra <u>valore misurato</u>  $\underline{c_t}$  e <u>valore vero</u> c(t + T/2) è:  $\frac{\Delta c}{c} = \frac{1 - e^{-\xi}}{\xi} - e^{-\xi/2}$ 

$$\frac{\Delta c}{c} = \frac{1 - e^{-\xi}}{\xi} - e^{-\xi/2}$$

avendo posto  $\xi = T/\tau$ .

Si vede quindi come nel semplice processo preso in esame l'errore percentuale dipenda dal rapporto  $\xi$  in un modo perfettamente determinabile e si può in tutta tranquillità estrapolare l'osservazione a situazioni più complesse per le quali le condizioni di errore dipenderanno dalla natura del processo (nel caso esaminato il parametro  $\tau$ ) e dal tempo di campionamento in modo più articolato e complesso di come appare in fig.3.4.

# 3.4. – Errori dovuti al campionamento

Gli errori da campionamento considerati nel paragrafo precedente sono stati valutati con riferimento ad un sistema particolarmente semplice ed è, quindi, ragionevole ritenere che nella pratica il campionamento produca effetti anche amplificati. Del resto, in un miscelatore reale la miscelazione non è mai completa e si deve sempre fare i conti con zone morte e cortocircuiti capaci di complicare notevolmente le formule usate per il diagramma di fig.3.4. Inoltre, anche senza effetti dovuti ai processi, il campionamento introduce motivi di imprecisione. Infatti, di solito i

campionamenti si eseguono sequenzialmente (le misure isolate nei problemi ambientali hanno solo funzioni esplorative) prelevando in uno dei modi seguenti:

- a) ad intervalli uguali (prelievi di ugual durata) e contigui come quando si preleva acqua consecutivamente con bottiglie di ugual capacità;
- b) ad intervalli uguali non contigui, ma regolari come quando si preleva acqua con bottiglie di ugual capacità ad intervalli di tempo regolari, ma non consecutivi;
- c) ad intevalli diseguali (prelievi di durata variabile) e contigui come quando si preleva acqua con bottiglie di capacità variabile, ma consecutivamente;
- d) ad intervalli diseguali non contigui e non regolari come quando si preleva acqua con bottiglie di capacità variabile ad intervalli di tempo non regolari e non consecutivi;

ed ognuno dei protocolli fornisce campioni diversi. Nella fig.3.5 sono illustrate le quattro tipologie campionarie in relazione ad una curva esattamente parabolica (ad esempio la qualità dell'acqua di

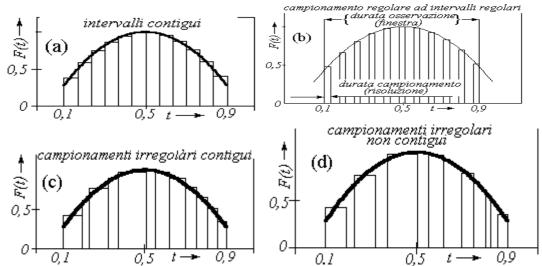


Fig. 3.5. Quattro diverse modalità di campionamento: (a) campionamenti di ugual durata ad intervalli di tempo regolari e contigui; (b) campionamenti di ugual durata ad intervalli di tempo regolari, ma non contigui; (c) campionamenti di durata variabile eseguiti ad intervalli di tempo irregolari, ma contigui; (d) campionamenti di durata variabile eseguiti ad intervalli di tempo irregolari e non contigui. La curva continua rappresenta il fenomeno.

un corpo idrico completamente miscelato nel quale la concentrazione di un tracciante varia nel

Fig. 3.6. Effetti del campionamento; confronto dei valori ottenuti con i campionamenti illustrati in fig. 10.

tempo secondo una tale legge). Le conseguenze sui dati sperimentali dovute alle quattro modalità di campionamento sono illustrate nella fig.3.6. La curva continua nei diagrammi di fig.3.5 viene vista dai quattro metodi come una funzione discreta rappresentata da punti dispersi attorno alla parabola che ha generato il processo. La presenza di dispersione dei dati che è solo conseguenza delle modalità di prelievo, può essere erroneamente ascritta a fenomeni inesistenti.

Altri effetti dovuti al campionamento che possono indurre errori di interpretazione sono quelli noti con il nome di *aliasing* (v. figg.3.7 e 3.8)<sup>(1)</sup>. La fig.3.7 mostra che, quando si campiona con

<sup>(1)</sup> L'effetto è quello stroboscopico che si osserva nei film quando viene inquadrata una ruota di una carrozza in rallentamento. Ad una particolare velocità della carrozza (quando la frequenza di rotazione della ruota è prossima

passo regolare T, un processo sinusoidale di frequenza  $v_1$  ci si può trovare con un fenomeno sinusoidale di frequenza  $v_2 = v_1 + 1/T$ . Infatti, nei confronti del campionamento scelto, i due processi appaiono indistinguibili. La fig.3.7 mette in evidenza che si ha *aliasing* solo quando  $T > 1/v_1$ , mentre la risoluzione con la quale si misurano i tempi  $T_r$  e la durata dell'esperimento (finestra di osservazione,  $T_f$ ) definiscono due frequenze estreme di campionamento: una  $(1/T_r)$ , che è la massima possibile perché la risoluzione dello strumento che misura il tempo non permette di osservare frequenze superiori; una  $(1/T_f)$ , che è la minima consentita dalla durata dell'esperimento.

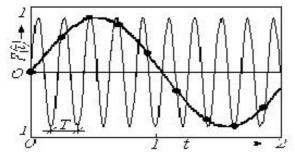


Fig. 3.7. – Effetto aliasing

Nei problemi reali i fenomeni di *aliasing* sono molto più frequenti di quanto si pensi (basta infatti ricordare che un qualsiasi segnale si può sempre trasformare in una somma di processi sinusoidali tramite sviluppo in serie di Fourier) e, purtroppo, sono anche sempre

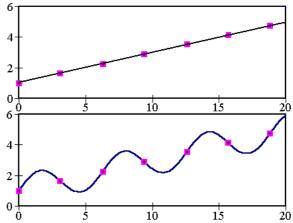


Fig. 3.8. Possibile errata interpretazione lineare di un processo ondulatorio osservato campionamento discreto.

difficili da individuare. Nel caso unidimensionale la loro natura è chiaramente descritta negli esempi trattati e nel caso bidimensionale l'effetto aliasing si manifesta come *effetto scacchiera* che si osserva quando si esagera con operazioni di *zoom* sulle immagini, ma nei processi multidimensionali le rappresentazioni grafiche non aiutano.

## 3.5. – Operazioni di misura

Le operazioni di misura, pur nella loro semplicità concettuale, presentano ancora aspetti non del tutto compresi, soprattutto in relazione alle scale delle grandezze da misurare. In campo ambientale un problema emblematico è quello delle coste (di un'isola, originariamente l'Inghilterra) o quello della rete di canali che drena un bacino imbrifero (una rete di segmenti che copre un'area!). In entrambi i casi esistono forme teoriche equivalenti (fiocco di neve di Koch e bacino di Peano) che risultano però irragionevoli agli occhi della geometria euclidea. Per questo non sembra superfluo riprendere i concetti primitivi dell'operazione di misura.

Disponendo di procedure per confrontare oggetti o, meglio, caratteristiche di oggetti (ad es., la lunghezza di un tavolo, il volume di una stanza, il tempo per andare da un posto ad un altro, la velocità di un'automobile, l'età di una persona, ecc.) è possibile classificare tutte le caratteristiche di tutti gli oggetti dell'universo come commensurabili o incommensurabili con una grandezza di riferimento (una caratteristica di un oggetto scelta come regolo (unità) di misura). Fra il regolo e le grandezze ad esso commensurabili si possono stabilire relazioni di copertura usualmente indicate come misura di quella caratteristica. Più precisamente si definisce come misura il *numero* di unità (e suoi multipli o sottomultipli) che occorrono per coprire la caratteristica dell'oggetto da misurare. Una tale operazione si rappresenta come rapporto fra la grandezza da misurare A e l'unità U:

$$\frac{A}{U} = m$$
 ovvero  $A = mU$ 

alla frequenza dei fotogrammi della ripresa) la rotazione della ruota sembra invertirsi perché i fotogrammi campionano la ruota regredendo.

dove m è il numero che rappresenta la misura di A rispetto ad U.

Si osserva che la grandezza fisica A è completamente indipendente dalla scelta dell'unità di misura U, la quale, per altro, è del tutto arbitraria. Segue che se si cambia l'unità dovrà cambiare anche la misura. Così se si riferisce A ad una nuova unità  $U_1$  è:

$$A = mU = m\frac{U}{U_1}U_1 = m_1U_1$$

dove:

 $m_1 = m \mu$ 

essendo:

$$\mu = \frac{U}{U_1} .$$

Il numero  $\mu$ , che esprime la misura di U rispetto ad  $U_1$ , viene detto fattore di conversione dall'unità U all'unità  $U_1$ .

Per facilitare le comunicazioni fra le persone si standardizzano le unità di misura, cioè se ne conviene una per ogni classe di grandezze omogenee formando così un *sistema di unità di misura*. Le unità di un sistema di misura non sono necessariamente tutte fissate in modo indipendente dato che fra diverse grandezze intercorrono relazioni (definizioni, principi e leggi fondamentali) universalmente riconosciute utilizzabili per legare tra loro unità non omogenee (ad esempio, velocità con spazio percorso e tempo impiegato a percorrerlo). Quanto detto consente di classificare le unità come:

- a) *fondamentali* se si fissano in modo indipendente, cioè, che non sono fra loro legate da alcuna relazione;
- b) derivate se si fissano a mezzo di relazioni derivandole da quelle fondamentali.

Si arriva quindi a definire i *sistemi di unità di misura* come gli insiemi di tutte e solo (*completezza*) le unità fondamentali necessarie ad una disciplina per esprimere ogni sua grandezza a mezzo di esse.

I problemi della misura sono emersi cercando di interpretare lunghezze di oggetti naturali, ma erano già apparsi in relazione ad alcune figure geometriche artificiali.

#### 3.6. – Confronto dei dati

I dati sperimentali e/o teorici vengono raccolti per essere messi in relazione con altri dati sperimentali e/o teorici. Ricordando che le relazioni presuppongono la commensurabilità è importante costruire algoritmi capaci di quantificare la "distanza" fra dati per poter dire "di quanto" due situazioni differiscano o meglio quando si possano considerare "uguali". La distanza d fra due dati  $\underline{x}_I$  e  $\underline{x}_2$  (intesi come vettori) è generalmente definita in senso euclideo dal quadrato  $d = (\underline{x}_I - \underline{x}_2)^2$ , ma questa non è l'unica possibilità di misurarla potendosi considerare relazioni molto più complesse  $(d = F(\underline{x}_I, \underline{x}_2))$  e fra queste il modulo e una qualsiasi delle potenze pari della differenza  $(\underline{x}_I - \underline{x}_2)$ . Il problema è infatti quello di ricondurre ad un numero o ad una figura semplice l'informazione da trattare. Ad esempio, si può avere la necessità di stabilire se una sorgente è più inquinante di un'altra mirando a proteggere la quantità di territorio interessata, indipendentemente dalla intensità dell'inquinamento. In questo caso si confrontano le aree inquinate dalle due sorgenti. Ma questa situazione assume significati più sottili se le due sorgenti sono una reale ed una virtuale, ovvero se si tratta di confrontare dati sperimentali e dati prodotti da un modello matematico del quale si vuole dare un giudizio di merito. Allora si può usare come criterio il rapporto  $\{(A_p \cap A_o)/(A_p \cup A_o)\}$  fra intersezione  $(\cap)$  e unione  $(\cup)$  delle aree individuate sperimentalmente  $(A_o)$  e teoricamente  $(A_p)$ .

Analogamente se si vuol effettuare un confronto di sovrapposizione temporale fra eventi (ad esempio nelle previsioni meteorologiche) si può usare il rapporto  $\{(T_v \cap T_o)/(T_v \cup T_o)\}$  fra intersezione e unione della persistenza dell'evento  $(T_p)$  e di quella  $(T_o)$  di riferimento.

In altre parole, non esiste un unico criterio di confronto fra eventi o fra dati e questa situazione rispecchia la percezione che ognuno di noi ha della bontà delle previsioni. Di fatto l'unico modo per confrontare oggetti, dati o eventi consiste nel convenire un criterio ed attenersi a quello fino a quando non se ne trovi uno che piace di più ricordando che è in ogni caso arbitrario.

### 3.7 – Confronti statistici

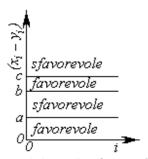


Fig.3.9. - Confronto di due insiemi di dati (x<sub>i</sub> e  $v_i$ ) e definizione delle favorevoli zone sfavorevoli al confronto.

decisione (v. fig.3.10):

Per effettuare un confronto è necessario (i) stabilire un criterio e (ii) prendere una decisione. Il criterio consiste nel definire una funzione da applicare ai dati da confrontare (ad esempio la loro differenza) ed uno o più intervalli del suo codominio favorevoli al confronto [ad esempio: in fig.3.9 gli intervalli  $(x_i - y_i) < a$  e  $b < (x_i - y_i) < c$ ]. La decisione consiste nel rigettare i dati sfavorevoli al confronto oppure nel considerarli con precauzione (ad esempio pesandoli diversamente).

In statistica si effettua una decisione definendo una proprietà e valutandone la presenza (*ipotesi nulla*,  $H_0$ ) o l'assenza (*ipotesi* alternativa,  $H_1$ ). La separazione delle due ipotesi si ottiene introducendo una funzione statistica (test statistico, X) che dipende dai dati disponibili e dalla proprietà utilizzata. L'accettazione o il rigetto dell'ipotesi nulla dipende dal fatto che il valore ottenuto per il test cada o meno in una regione critica specificata in anticipo.

L'applicazione di un test può portare a due tipi di errori di

- a) ERRORE DEL I° TIPO se si rigettano ipotesi vere;
- b) ERRORE DEL II° TIPO se si accettano ipotesi false.

Gli errori del I° tipo sono caratterizzati da una probabilità  $\alpha$  (nota la distribuzione di X, ad

ogni valore X' di X corrisponde una probabilità  $\alpha'$ ) che può superare o meno un valore prefissato. Il valore  $\alpha_0$ assunto come limite per accettare o respingere l'ipotesi  $H_{\theta}$  viene detto *livello di significatività* del test. Il valore di  $\alpha$  non basta per una decisione coerente con le ipotesi perchè non informa sugli errori del IIº tipo. Questi vengono caratterizzati da una probabilità  $\beta$ . La Fig.3.10. - Tabella delle decisioni probabilità  $(1 - \beta)$  è la potenza del test rispetto statistiche.

DECISIONE	ACCETTA IPOTESI H.	RIFIUTA IPOTESI H。
VERA	NESSUN ERRORE	ERRORE DEL I° TIPO probabilità = ¤
FALSA	ERRORE DEL II° TIPO probabilità = β	NESSUN ERRORE

all'ipotesi alternativa e misura la probabilità di non commettere errori del II° tipo intendendo errate le situazioni che ricadono nell'ipotesi alternativa.

È chiaro che il dominio delle ipotesi alternative da contrapporre all'ipotesi nulla condiziona l'ampiezza dello spazio degli eventi possibili e quindi anche la potenza selettiva del test<sup>(2)</sup>. Naturalmente, limitando il campo delle ipotesi alternative si realizzano test più potenti, ma se ne riduce la generalità.

<sup>(2)</sup> Da una variabile casuale Z si estrae un campione  $U = (Z_1, Z_2, ..., Z_n)$  le cui variabili casuali  $Z_1, Z_2, ..., Z_n$ , sono tutte identiche alla variabile casuale di partenza Z ed indipendenti le une dalle altre come nello schema di Bernoulli dell'estrazione da un'urna con ripristino della composizione. Verificare un'ipotesi statistica,  $H_0$ , significa ripartire lo spazio campionario a n dimensioni W (costituito da tutti i possibili valori del vettore U), in due regioni complementari,  $\hat{w}$  ed  $A = \hat{W} - w$ . L'ipotesi  $H_0$  viene rifiutata se il campione osservato u cade nella regione w (detta *regione critica*) ed accettata se u cade in A (detta regione di accettazione).

Non è banale segnalare che quando si effettua il confronto di dati, ad esempio per decidere fra due modelli quale sia il migliore, si deve evitare di cercare risposte superflue, dimenticando, ad esempio, che un polinomio di grado n è un caso particolare dell'analogo polinomio di grado n+1 oppure che l'aggiunta di un parametro può non migliorare un *fitting*. Su questa linea si osserva anche che i confronti fra grandezze incommensurabili sono privi di significato fisico e che quindi si deve porre particolare attenzione agli indicatori che si confrontano ed alla scelta dell'ipotesi  $H_0$ .

TEST { parametrico: la pdf del test dipende dalla pdf dei dati potesi a confronto «semplici» se sono noti tutti i parametri della pdf dei dati - ipotesi a confronto «composite» se è indeterminato almeno un parametro non parametrico: la pdf del test non dipende dalla pdf dei dati

Fig. 3.11. - Schema di classificazione dei test sulla base della distribuzione di probabilità dei dati e dei test.

La bontà di un test consiste nel rendere minimi gli errori sia del  $I^{\circ}$  che del  $II^{\circ}$  tipo, tuttavia, il fatto che  $\alpha_0$  sia generalmente prefissato induce a considerare migliore il

test con potenza massima. Inoltre, ad un buon test si richiede che sia privo di errori sistematici (unbiased) rispetto all'ipotesi alternativa, ciò che si realizza quando  $\alpha_0 \le (1 - \beta)$ .

Per completare il quadro panoramico relativo ai *test di ipotesi* se ne ricorda una classificazione sulla base delle funzioni di distribuzione della probabilità (*pdf*) dei dati e dei *test* (v. fig.3.10). In particolare lo schema di fig.3.11 illustra la differenza fra i *test* parametrici e quelli non parametrici.

### - APPENDICE -OSCILLAZIONI SMORZATE

Uno strumento di misura generale è costituito da un indicatore che scorre su una scala e in assenza

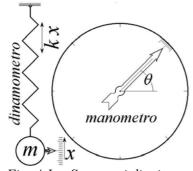


Fig. A.I. – Strumenti di misura.

di sollecitazioni ritorna in una posizione di riposo, lo zero, e per raggiungere il valore da misurare deve vincere delle resistenze. Può essere visto come un dinamometro o un manometro o altro il cui indicatore (una lancetta) si muove per effetto della "forza" da misurare che è bilanciata dall'azione di richiamo. L'indicatore si muove per raggiungere una posizione di equilibrio, ma il moto è contrastato da un'inerzia e da resistenze dissipative. L'equivalente meccanico di un tale apparecchio è rappresentato in fig.A.I dove non è stato disegnato il richiamo della lancetta manometrica perché nella pratica viene realizzato in modi molto vari.

Considerando resistenze proporzionali alla velocità dell'indicatore (attriti), l'equazione che descrive il moto di avvicinamento alla posizione di equilibrio (valore da misurare)<sup>(3)</sup> è:

$$\frac{inerzia}{m \frac{d^2 x}{dt^2} + c \frac{dx}{dt}} + kx = forze di richiamo valore da misurare$$
[A.1]

L'equazione [A.1] ha come integrale generale l'integrale generale dell'equazione differenziale omogenea ad essa associata più un integrale particolare che, come è facile verificare, è: x=f/k=h.

<sup>(3)</sup> Si considerano spostamenti rettilinei come nel caso del dinamometro, ma nulla cambia nel caso di spostamenti angolari come nel caso dei manometri.

L'equazione caratteristica dell'equazione omogenea corrispondente ad [A.1] è  $my^2+cy+k=0$  le cui radici sono:

$$y = \frac{-c \pm \sqrt{c^2 - 4mk}}{2m} = \frac{-c \pm \sqrt{\Delta}}{2m}$$

Il termine  $\sqrt{\Delta}$  prospetta tre casi (i=1, 2, 3) distinti corrispondenti a  $\Delta_1=0, \Delta_2>0$  e  $\Delta_3<0$ , cosicché, posto:

$$\gamma_i = -\frac{c_i}{2m_i} \text{ (con } i = 1, 2, 3), \qquad \beta = \frac{\sqrt{\Delta_2}}{2m_2} \qquad \text{e} \qquad \omega = \frac{\sqrt{-\Delta_3}}{2m_3},$$

in corrispondenza ai tre casi, l'integrale generale<sup>(4)</sup> dell'equazione [A.1] risulta (con  $a_1$ ,  $a_2$ ,  $a_3$ ,  $b_1$ ,  $b_2$ ,  $b_3$  costanti di integrazione per i tre casi):

$$x_1 = (h - h_0)e^{\gamma_1 t} \{a_1 t + b_1\} + h$$
 per  $\Delta = 0$  [A.2]

$$x_2 = (h - h_0)e^{\gamma_2 t} \left\{ a_2 + b_2 e^{(\beta - \gamma_2)t} \right\} + h$$
 per  $\Delta > 0$  [A.3]

$$x_3 = (h - h_0)e^{\gamma_3 t} \{a_3 \cos \omega t + b_3 \sin \omega t\} + h$$
 per  $\Delta < 0$ , [A.4]

Le cui derivate prime sono:

$$\frac{dx_{1}}{dt} = (h - h_{0})e^{\gamma_{1}t} \{a_{1} + \gamma_{1}[a_{1}t + b_{1}]\}$$
 per  $\Delta = 0$ 

$$\frac{dx_2}{dt} = (h - h_0)e^{\gamma_2 t} \left\{ \gamma_2 a_2 + \beta b_2 e^{(-\gamma_2 + \beta) t} \right\}$$
 per  $\Delta > 0$ 

$$\frac{dx_3}{dt} = (h - h_0)e^{\gamma_3 t} \{ \gamma_3 [a_3 \cos \omega t + b_3 \sin \omega t] + \omega [b_3 \cos \omega t - a_3 \sin \omega t] \}$$
 per  $\Delta < 0$ .

Per quanto riguarda le condizioni al contorno si considera che inizialmente l'indicatore x è fermo al valore  $h_0$ , cioè  $x(0)=h_0$  e  $dx/dt|_{t=0}=0$  e si ottiene la tabella.

Tabella – Costanti di integrazione per i tre casi.

_ tubella Costanti ai integrazione per i tre cast.			
$\Delta_i$	$a_i$	$b_i$	
$c_1^2 - 4m_1k_1 = 0$	$-\frac{c_1}{2m_1}$	-1	
$c_2^2 - 4m_2k_2 > 0$	$\frac{-\sqrt{c_2^2 - 4m_2k_2}}{c_2 + \sqrt{c_2^2 - 4m_2k_2}}$	$\frac{-c_2}{c_2 + \sqrt{c_2^2 - 4m_2k_2}}$	
$c_3^2 - 4m_3 k_3 < 0$	-1	$-\frac{c_3}{\sqrt{4m_3k_3-c_3^2}}$	

In definitiva si ha:

$$x_{1} = (h - h_{0})e^{-\frac{c_{1}}{2m_{1}}t} \left\{ -1 - \frac{c_{1}}{2m_{1}}t \right\} + h$$

$$x_{2} = (h - h_{0})e^{-\frac{c_{2}}{2m_{2}}t} \left\{ -\frac{\sqrt{c_{2}^{2} - 4m_{2}k_{2}}}{c_{2} + \sqrt{c_{2}^{2} - 4m_{2}k_{2}}} - \frac{c_{2}}{c_{2} + \sqrt{c_{2}^{2} - 4m_{2}k_{2}}}e^{\frac{c_{2} + \sqrt{c_{2}^{2} - 4m_{2}k_{2}}}{2m_{2}}t} \right\} + h$$
[A.6]

<sup>(4)</sup> Per stabilire che le funzioni indicate sono effettivamente gli integrali generali dei rispettivi casi, basta verificare che soddisfano l'equazione [A.1] con valori arbitrari delle costanti di integrazione perché il teorema di unicità degli integrali garantisce che quelle funzioni sono proprio le soluzioni cercate.

$$x_{3} = (h - h_{0})e^{-\frac{c_{3}}{2m_{3}}t} \left\{ -\cos\left(\frac{\sqrt{4mk_{3} - c_{3}^{2}}}{2m_{3}}t\right) - \frac{c_{3}}{\sqrt{4mk_{3} - c_{3}^{2}}} \operatorname{sen}\left(\frac{\sqrt{4mk_{3} - c_{3}^{2}}}{2m_{3}}t\right) \right\} + h$$
 [A.7]

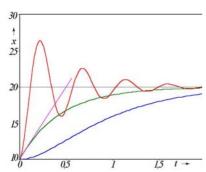


Fig. A.II. – Smorzamento critico (in verde), subcritico (in bleu) e oscillante (in rosso). La retta in colore fucsia ha il coefficiente angolare  $1/\tau = c/2m$ .

La  $x_3(t)$  (eq. A.7) ha andamento oscillante smorzato, mentre le altre due tendono asintoticamente al valore di misura, h. La  $x_1(t)$  (eq. A.5) tende ad h più rapidamente della  $x_2(t)$  ed è la curva di transizione da asintotica ad oscillante. Lo *smorzamento critico* (quello della  $x_1(t)$ ) segue una curva che tende al valore da misurare più velocemente delle altre, ma richiede comunque un tempo infinito per raggiungerlo per cui, nella pratica, conviene ricorrere a oscillazioni smorzate che, sorpassando alcune volte (3 o 4) il valore di misura, rendono le letture più veloci e più precise (v. fig. A.II, costruita con  $c_1/m_1 = c_2/m_2 = c_3/m_3 = c/m$ ). Si osserva anche che se c=0 (assenza di attriti) è  $\Delta < 0$  e l'indicatore oscilla indefinitamente senza smorzamento. Inoltre, lo smorzamento critico individua un tempo caratteristico dello strumento (detto tempo di salita)  $\tau = 2m/c$  per il quale:

$$x(\tau) = h - (h - h_o) \ 2/e = h_o + (h - h_o) \ \{1 - \}2/e \approx h_o + 0.3 \ (h - h_o).$$

In altri termini  $\tau$  rappresenta il tempo necessario allo strumento per raggiungere il 30% del valore da misurare. Questa frazione aumenta se, a parità di 2m/c, il valore di k è tale da rendere il sistema oscillante.