

# COMPLEMENTI DI STATISTICA <sup>1</sup>

Bruno Crippa <sup>(1)</sup> – Luigi Mussio <sup>(2)</sup>

<sup>(1)</sup> Università degli Studi di Milano – Dipartimento di Scienze della Terra "Ardito Desio" – Via Cicognara, 7 – 20129 Milano  
Tel. 02-503-18474 – Fax 02-503-18489 – e-mail:bruno.crippa@unimi.it

<sup>(2)</sup> Politecnico di Milano – DICA – Piazza Leonardo da Vinci, 32 – 20133 Milano  
Tel. 02-2399-6501 – Fax 02-2399-6602 – e-mail luigi.mussio@polimi.it

## RIASSUNTO

Dopo la statistica descrittiva, l'inferenza statistica e la teoria della stima, se del caso, integrate con quei dettagli che ne arricchiscono la trattazione, ad esempio, come il campionamento, l'analisi di varianza e la regressione multipla, differenti possono essere molte altre tematiche complementari. A riguardo, i processi stocastici permettono di studiare la natura fine dei dati delle osservazioni e, in questo stesso ambito, sono poi presentate anche alcune estensioni della distribuzione di Poisson. Il controllo statistico della qualità, la teoria della robustezza, alcuni esempi di analisi multivariata ed una riflessione sulla storia, la teoria e la pratica di un metodo scientifico sono quattro appendici che concludono questo lavoro.

## PARTE I – PROCESSI STOCASTICI

### 1.1 Introduzione

In un esempio di un particolare processo stocastico, ogni spezzata  $x_1(t)$ ,  $x_2(t)$ , ecc. si può considerare come un valore campionario del processo, ottenuto diagrammando, in funzione del tempo, il numero casuale di arrivi dell'evento considerato che, a sua volta, segue la distribuzione di Poisson.

Altre serie di osservazioni conducono alla determinazione di altre *realizzazioni* del processo  $x_r$  ( $r = 1, 2, \dots, n, \dots, \infty$ ), esattamente come avviene nelle variabili casuali all'estrazione di nuovi valori campionari.

In generale, un processo stocastico può essere descritto come una funzione di due variabili  $X(x, t)$ , tali che esso possa assumere una delle seguenti quattro interpretazioni, a seconda di quale variabile si fissa:

- se è fissato  $r = r_0$ , si individua la realizzazione, funzione del solo tempo;
- se si fissa anche  $t = t_0$ , fissato  $r = r_0$ , si ottiene un numero, cioè il valore:  $x = X(t_0)$ , presentato dalla realizzazione  $r_0$  nell'istante  $t_0$  (come mostra la Figura 1.1.1);
- se è fissato  $t = t_0$ , lasciando  $r$  variabile, si ha l'insieme dei valori  $x$ , assunti dalle varie realizzazioni all'istante  $t_0$  che, per poter definire il processo stocastico, devono presentarsi sotto forma di variabili casuali, cioè associati con una funzione densità di probabilità: ad esempio, nel processo di Poisson:

$$p_{(x_r)} = e^{-(\lambda t_0)} (\lambda t_0)^{x_r} / x_r! \quad \text{con} \quad r = 1, 2, \dots, \infty$$

---

<sup>1</sup> La prima parte di questo lavoro riporta, pressoché integralmente, quanto esposto nei capitoli 10 ed 11 del libro: Fondamenti di statistica, di Giovanna Togliatti (Hoepli, Milano, 1976), dove le note, scritte dagli autori del presente lavoro, servono a colmare i quasi quaranta anni passati dall'epoca di edizione del libro suddetto, così da rendere questi due capitoli ancora pienamente attuali.

□ se sia  $t$  che  $r$  sono lasciati liberi di variare, la notazione  $X(t, x_p)$  rappresenta il processo stocastico, nel suo insieme, tuttavia indicato in seguito, per semplificare i simboli, solo come  $X(t)$ , sottintendendo così la presenza delle possibili infinite realizzazioni.

Quasi sempre la funzione del tempo, rappresentante la singola realizzazione, è una curva irregolare, non esprimibile con formula analitica. Un esempio è dato dal percorso, compiuto da una particella, in conseguenza degli scontri con le particelle, vicine ad essa (come nei moti Browniani). Le varie realizzazioni sono collegate ai percorsi delle singole particelle e, pur conoscendone il valore  $x_p(t)$ , fino a  $t_0$ , non si può predirne l'andamento per  $t > t_0$ .

Per contro, si possono avere altri tipi di processi, ad esempio, come quello che descrive la forza elettromotrice di una famiglia di generatori in cui l'ampiezza, la frequenza e la fase sono variabili casuali, dove una particolare terna  $a(r)$ ,  $\omega(r)$  e  $\varphi(r)$  dipende dal generatore considerato. Il processo stocastico:  $X(t) = a \sin(\omega t + \varphi)$  è costituito da un insieme di sinusoidi e, di ogni realizzazione, si può anche predire l'andamento, se lo si conosce per  $t < t_0$ .

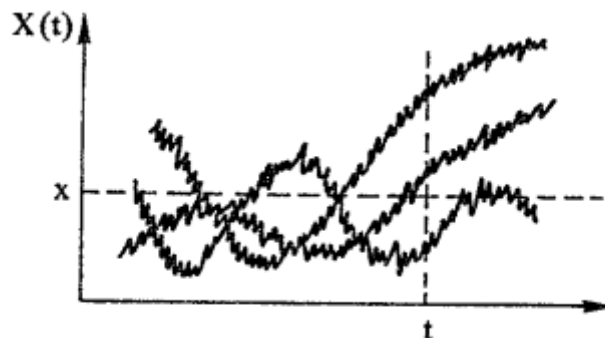


Fig. 1.1.1 – Insieme di realizzazioni costituenti un processo stocastico

Come già detto, per uno specificato istante  $t$ , il processo  $X(t)$  è una variabile casuale la cui distribuzione, di solito, dipende da  $t$ . Indicando questa funzione distribuzione con  $F(x, t)$ , si ha:

$$F(x, t) = P[X(t) \leq x]$$

Questo significa che  $F(x, t)$  è la probabilità che, al tempo  $t$ , le funzioni, costituenti il processo stocastico, non superino  $x$ . La funzione  $F(x, t)$  è detta distribuzione del primo ordine e la corrispondente funzione densità di probabilità è:

$$f(x, t) = \frac{\partial F(x, t)}{\partial x}$$

Invece se si considerano in due tempi diversi le variabili casuali  $X(t_1)$  e  $X(t_2)$ , si può definire la distribuzione del secondo ordine, del processo  $X(t)$ , come distribuzione bidimensionale, in genere, dipendente da  $t_1$  e  $t_2$ :

$$F(x_1, x_2, t_1, t_2) = P((X(t_1) \leq x_1) \cap (X(t_2) \leq x_2))$$

e la corrispondente densità di probabilità bidimensionale:

$$f(x_1, x_2, t_1, t_2) = \frac{\partial^2 F(x_1, x_2, t_1, t_2)}{\partial x_1 \partial x_2}$$

Un processo stocastico è statisticamente determinato, quando si conoscono tutte le sue distribuzioni di ordine  $n$ :

$$F(x_1, x_2, \dots, x_n, t_1, t_2, \dots, t_n) = P((X(t_1) \leq x_1) \cap (X(t_2) \leq x_2) \cap \dots \cap (X(t_n) \leq x_n))$$

per qualsiasi  $n$  e  $t_1, t_2, \dots, t_n$ , da cui si ottiene la densità di probabilità di ordine  $n$ , derivando rispetto a tutte le variabili  $x_1, x_2, \dots, x_n$ ,

## 1.2. Momenti d'insieme: media, auto-correlazione, auto-covarianza, cross-correlazione e covarianza

Di solito, la media d'insieme di un processo stocastico  $X(t)$  è funzione del tempo e, ad ogni istante  $t$ , è data dalla media della variabile casuale che il processo origina in  $t$ :

$$\mu(t) = M(X(t)) = \int_{-\infty}^{+\infty} x f(x, t) dx \quad (2.1)$$

Invece l'auto-correlazione di un processo stocastico dipende da  $t_1$  e  $t_2$ , ed è definita come:

$$R_{XX}(t_1, t_2) = M(X(t_1)X(t_2)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (2.2)$$

Si definisce ulteriormente l'auto-covarianza, ancora dipendente da  $t_1$  e  $t_2$ :

$$C_{XX}(t_1, t_2) = M((X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))) = R_{XX}(t_1, t_2) - \mu(t_1)\mu(t_2) \quad (2.3)$$

Un caso particolare della (2.3) è la varianza del processo:

$$\sigma_{X(t)}^2 = C_{XX}(t, t) = R_{XX}(t, t) - \mu^2(t) \quad (2.4)$$

Dati due processi stocastici  $X(t)$  e  $Y(t)$ , si definiscono anche la cross-correlazione e la cross-covarianza:

$$R_{XX}(t_1, t_2) = M(X(t_1)Y(t_2)) \quad (2.5)$$

$$C_{XY}(t_1, t_2) = R_{XY}(t_1, t_2) - \mu_X(t_1)\mu_Y(t_2) \quad (2.6)$$

Se il processo è complesso, le (2.2) e (2.3) si modificano come segue:

$$R_{XX} = M(X(t_1)X^*(t_2)) \quad C_{XX} = R_{XX} - \mu(t_1)\mu^*(t_2)$$

dove con il simbolo \* si sono indicati i valori coniugati (analogamente si modificano le (2.5) e (2.6)). Infine il momento  $M(X(t_1) \cdot X(t_2) \cdot \dots \cdot X(t_n))$  è la statistica di ordine  $n$  del processo stocastico e dipende da  $t_1, t_2, \dots, t_n$ .

### 1.3 Stazionarietà

Un processo stocastico si dice stazionario in senso stretto, quando le sue statistiche, di tutti gli ordini, non cambiano, in conseguenza di un cambiamento di origine dei tempi. Questo significa che i due processi stocastici  $X(t)$  e  $X(t+t_0)$  hanno le stesse statistiche per ogni  $t_0$ . Due processi stocastici si dicono congiuntamente stazionari in senso stretto, se le statistiche congiunte di  $X(t)$  e  $Y(t)$  sono le stesse di quelle di  $X(t+t_0)$  e  $Y(t+t_0)$ , per ogni  $t_0$ .

Dalla definizione, segue che, se un processo stocastico è stazionario, la sua media è costante nel tempo e la sua auto-correlazione ed auto-covarianza dipendono solo da:  $\tau = t_2 - t_1$ :

$$M(X(t)) = \mu = \text{cost} \quad (3.1)$$

$$R_{XX}(\tau) = M(X(t+\tau)X(t)) = R_{XX}(-\tau) \quad (3.2)$$

ed analogamente per due processi stocastici congiuntamente stazionari:

$$R_{XY}(\tau) = M(X(t+\tau)Y(t)) = R_{YX}(\tau) \quad (3.3)$$

Oltre alla stazionarietà in senso stretto, esistono molti tipi di stazionarietà, di ordine  $k$ , asintotica in un intervallo, periodica, ecc., a seconda del tipo di statistica per la quale esiste l'invarianza rispetto a traslazioni nel tempo. Una delle più importanti forme di questa stazionarietà parziale è quella *in senso lato*, secondo la quale si richiede soltanto che la media sia una costante e l'auto-correlazione dipenda solo da:  $\tau = t_2 - t_1$ .

Se poi il processo stocastico è costituito da variabili casuali tutte normali, esso è compiutamente definito dai momenti di primo e secondo ordine e la stazionarietà in senso lato implica anche quella in senso stretto. Infine due processi stocastici si dicono congiuntamente stazionari in senso lato, se la loro cross-correlazione dipende solo da:  $\tau = t_2 - t_1$ .

#### 1.4. Momenti temporali

Dato un processo stocastico reale  $X(t)$  si possono definire, sotto condizioni molto vaste, i seguenti integrali:

$$m = \int_a^b X(t) dt$$

$$m^2 = \int_a^b X(t_1) dt_1 \int_a^b X(t_2) dt_2 = \int_a^b \int_a^b X(t_1) X(t_2) dt_1 dt_2$$

Potendosi scambiare liberamente l'operatore  $M$  (media) con il segno d'integrale, si determinano la media e la varianza di  $s$ , a sua volta rappresentante la variabile casuale delle aree sotto le varie realizzazioni del processo stocastico, fra  $t = a$  e  $t = b$ :

$$M(m) = M\left(\int_a^b X(t) dt\right) = \int_a^b M[X(t)] dt = \int_a^b \mu(t) dt$$

$$\sigma_m^2 = M(m^2) - M^2(m) = \int_a^b \int_a^b (R(t_1, t_2) - \mu(t_1)\mu(t_2)) dt_1 dt_2 = \int_a^b \int_a^b C(t_1, t_2) dt_1 dt_2 \quad (4.1)$$

Supposto che il processo stocastico  $X(t)$  sia stazionario, si può definire una variabile casuale, su questo, analoga a  $m$ , così definita:

$$s = \frac{1}{2T} \int_{-T}^T X(t) dt$$

cosicché introducendo:  $\tau = t_2 - t_1$ , se il processo è stazionario, le (4.1) diventano:

$$M(s) = \mu \quad (4.2)$$

$$\sigma_s^2 = \frac{1}{4T^2} \int_{-T}^T \int_{-T}^T C(t_2 - t_1) dt_1 dt_2 = \frac{1}{T} \int_0^{2T} \left(1 - \frac{\tau}{2T}\right) C(\tau) d\tau = \frac{1}{T} \int_0^{2T} \left(1 - \frac{\tau}{2T}\right) (R(\tau) - \mu^2) d\tau \quad (4.3)$$

Questa premessa permette di introdurre il concetto di momento temporale di un processo stocastico e, in particolare, di media ed auto-correlazione temporale. Di conseguenza, si considerino i limiti:

$$\bar{x} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T X(t) dt \quad (4.4)$$

$$r = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T X(t + \tau) X(t) dt \quad (4.5)$$

E' ovvio che  $\bar{x}$  e  $r$  media e auto-correlazione temporali, sono variabili casuali e, in generale, non possono essere sostituite alla media ed all'auto-correlazione d'insieme:  $M(X(t))$  e  $M(X(t + \tau)X(t))$ . Tuttavia se valgono le seguenti relazioni:

$$M(\bar{x}) = M(X(t)) = \mu \quad \text{con} \quad \sigma_{\bar{x}}^2 = 0 \quad (4.6)$$

$$M(r) = M(X(t + \tau)X(t)) = R_{XX}(\tau) \quad \text{con} \quad \sigma_r^2 = 0 \quad (4.7)$$

cioè se la media temporale e l'auto-correlazione temporale convergono, in probabilità, rispettivamente a  $\mu$  e  $R_{XX}(\tau)$ , allora le espressioni (4.4) e (4.5) possono essere utilizzate per definire la media e l'auto-correlazione di un processo stocastico, in sostituzione dei corrispondenti momenti d'insieme, in quanto essi sono rispettivamente uguali con probabilità uno.

### 1.5 Ergodicità

Uno dei principali problemi, per le variabili casuali, riguarda la possibilità (o meno) di stimare le loro statistiche, a partire da uno o pochi elementi campionari. Nel campo dei processi stocastici, si presenta un problema del tutto analogo in quanto, invece delle infinite realizzazioni del processo, se ne posseggono, di solito, una o poche. Pertanto si tratta di definire, sotto quali condizioni è possibile stimare le statistiche del processo stocastico, in base a quelle di una singola realizzazione.

Se questo è possibile, cioè se, con probabilità uno, tutte le statistiche del processo stocastico si possono stimare da quelle di una singola funzione  $X(t, \xi)$ , il processo si dice *ergodico* in senso stretto e, poiché le statistiche di una singola realizzazione si possono ricavare solo come momenti temporali, l'ergodicità si definisce anche come uguaglianza fra i momenti temporali ed i momenti d'insieme. Esistono tuttavia forme di ergodicità più limitate e più comuni, ad esempio, dove vale solo la (4.6), oppure la (4.6) e (4.7), in questo caso, si dice che il processo è ergodico nella media o nell'auto-correlazione.

E' evidente l'analogia fra la media campionaria  $\bar{x}$  e la media temporale. Infatti se si osserva una porzione di una singola realizzazione di un processo stocastico e si esegue la media temporale, si ottiene un numero  $\bar{x}_T$  che appartiene ad una variabile casuale, proprio come appartiene la media campionaria di una variabile casuale. Allora se si usa  $\bar{x}_T$  come stima di  $M(X(t))$ . la sua attendibilità è data dalla varianza di  $\bar{x}_T$ , potrà applicando la disuguaglianza di Tchebyceff, per ricavare la probabilità che l'errore commesso sia inferiore a  $\varepsilon$  assegnato:

$$P(|\bar{x}_T - M(X(t))| < \varepsilon) \geq 1 - \sigma_{\bar{x}_T}^2 / \varepsilon^2$$

### Esempio 1.5.1

Un processo stocastico  $X(t)$ , tale che:

$$M(X(t)) = 0 \quad \text{e} \quad R_{XX}(\tau) = e^{-2\lambda|\tau|}$$

è ergodico nella media. A questo scopo, bisogna controllare che la media delle medie temporali sia uguale alla media d'insieme e che la varianza delle medie temporali tenda a zero, per  $T \rightarrow \infty$ :

$$M(\bar{x}) = M\left(\frac{1}{2T} \int_{-T}^T X(t) dt\right) = \frac{1}{2T} \int_{-T}^T M(X(t)) dt = 0$$
$$\sigma_{\bar{x}}^2 = \frac{1}{T} \int_0^{2T} \left(1 - \frac{\tau}{2T}\right) e^{-2\lambda|\tau|} d\tau = \frac{1}{2\lambda T} - \frac{1}{8\lambda^2 T^2} (1 - e^{-4\lambda T}) \quad \text{da cui} \quad \lim_{T \rightarrow \infty} \sigma_{\bar{x}}^2 = 0$$

e dato che entrambe le condizioni sono verificate, il processo è ergodico nella media.

In modo analogo, si può controllare l'ergodicità dell'auto-correlazione, ma è più complicato perché, mentre per controllare l'ergodicità della media basta la conoscenza delle statistiche di primo e secondo ordine, per controllare l'ergodicità dell'auto-correlazione, occorrono anche quelle di terzo e quarto ordine.

### 1.6 Autocorrelazione e spettri di potenza

Per processi stocastici reali e stazionari si ha:

$$R_{XX}(\tau) = M(X(t_1)X(t_2)) = M(X(t_1 - t_2)X(0)) = M(X(0)X(t_2 - t_1)) = R_{XX}(\tau) = R_{XX}(-\tau) \quad (6.1)$$

Questo significa che l'auto-correlazione è una funzione pari, cioè simmetrica rispetto all'origine  $\tau = 0$ , e, si può anche dimostrare che, sempre nell'origine, assume un valore positivo (o nullo) ed ha il suo massimo. Infatti dalla (6.1) si ha:

$$R(0) = M((X(t))^2) \geq 0$$

inoltre:

$$M((X(t+\tau) \pm X(t))^2) = M((X(t+\tau))^2 + (X(t))^2 \pm 2X(t+\tau)X(t)) = 2(R(0) \pm R(\tau)) \geq 0$$

da cui:

$$|R(\tau)| \leq R(0)$$

Forme tipiche di autocorrelazione sono riportate nella Figura 1.6.1.

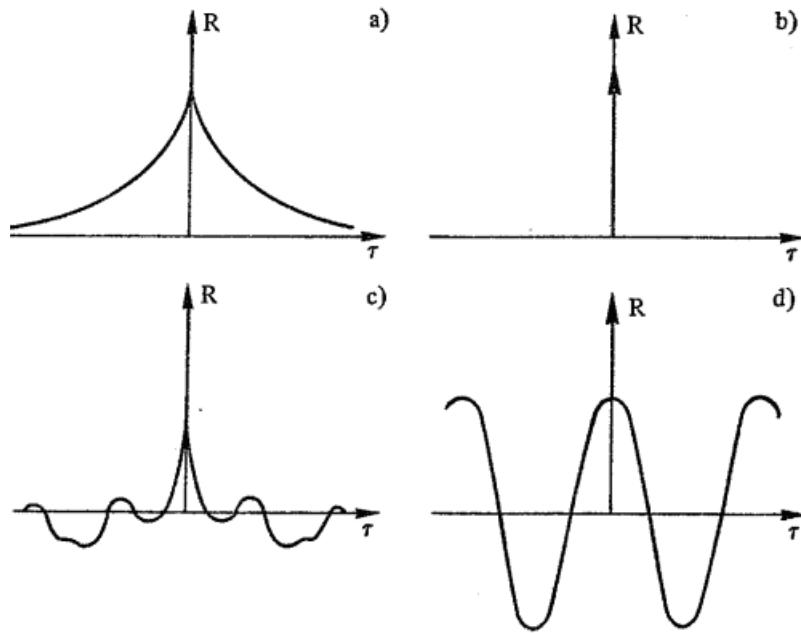


Fig. 1.6.1 – Diversi tipi di auto-correlogrammi

- ❑ La Figura 1.6.1.a è tipica di un processo stocastico *casuale*. Il rapido decadimento dell'autocorrelazione esprime la scarsa dipendenza fra lo stato del processo al tempo  $t$  e quello al tempo  $t + \tau$ .
- ❑ Nel caso di un processo stocastico completamente casuale, come per il *rumore bianco*, la funzione di autocorrelazione assume la forma della Figura 1.6.1.b, dove si ha un impulso uguale alla varianza del processo stocastico, per  $\tau = 0$ , e zero altrove. Tuttavia occorre notare che il *rumore bianco* è un fenomeno puramente teorico la cui natura si intuisce proprio dalla forma dell'auto-correlazione, in pratica, realizzato con una buona approssimazione, tramite apparecchiature elettroniche, capaci di simulare tutte le infinite possibili frequenze.
- ❑ La Figura 1.6.1.c è tipica di un processo in cui si sommano alcune componenti oscillatorie ed una componente casuale.
- ❑ La Figura 11.6.1.d è nuovamente un caso limite, essendo infatti l'auto-correlazione di una sinusoide la cui forma è perfettamente riprodotta. Nell'analisi di un segnale, la sola ispezione della forma dell'auto-correlazione rivela la presenza o meno di fenomeni oscillatori (cui, di solito, è sovrapposto un rumore) e che sono evidenziati da oscillazioni, più o meno smorzate, nell'andamento dell'auto-correlogramma (come è detta la rappresentazione grafica della funzione di auto-correlazione).

Si pone allora il problema dell'individuazione della o delle frequenze di questa parte oscillatoria del processo stocastico, che si può anche chiamare parte sistematica, in contrapposizione con la parte casuale, presente nel processo stocastico stesso. A questo scopo, come presentato in seguito, si presta bene il cosiddetto *spettro di potenza* (o densità spettrale) del processo stocastico:

$$S(\omega) = \int_{-\infty}^{+\infty} R(\tau) e^{-j\omega\tau} d\tau \quad (6.2)$$

che è la trasformata di Fourier della funzione di auto-correlazione.

Se il processo  $X(t)$  è reale,  $R(\tau)$  è reale e pari, e così è anche  $S(\omega)$ , per cui la (6.2) si può scrivere:

$$S(\omega) = \int_{-\infty}^{+\infty} R(\tau) \cos(\omega\tau) d\tau \quad (6.3)$$

valendo ovviamente la formula inversa:

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(\omega) \cos(\omega\tau) d\omega \quad (6.4)$$

Per  $\tau = 0$ , la (6.4) assume il valore:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} S(\omega) d\omega = R(0) = M((X(t))^2) \geq 0 \quad (6.5)$$

che esprime l'area totale sotto la funzione  $S(\omega)/2\pi$ : non negativa ed equivalente alla potenza media del processo stocastico. Si può poi ulteriormente dimostrare che risulta  $S(\omega) \geq 0$  per qualsiasi  $\omega$ .

Analogamente si definisce un cross-spettro di potenza di due processi  $X(t)$  e  $Y(t)$  come la trasformata di Fourier della loro cross-correlazione:

$$S_{XY}(\omega) = \int_{-\infty}^{+\infty} R_{XY}(\tau) e^{-j\omega\tau} d\tau \quad (6.6)$$

$$R_{XY}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{XY}(\omega) e^{-j\omega\tau} d\omega \quad (6.7)$$

Per  $\tau = 0$ , la (6.7) assume il valore:

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{XY}(\omega) d\omega = R_{XY}(0) = M(X(t)Y(t)) \quad (6.8)$$

Ad esempio, se  $X(t)$  è la tensione, ai due terminali di una presa, e  $Y(t)$  è la corrente d'entrata, in un apparecchio, l'area sotto la curva  $S_{XY}/2\pi$  uguaglia la potenza media dell'apparecchio.

La definizione di spettro di potenza, formulata con la (6.2), è spesso sostituita dalla seguente espressione:

$$S(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2T} \left( \int_{-T}^{+T} X(t) e^{-j\omega t} dt \right)^2$$

Tuttavia perché la sostituzione dello spettro di potenza temporale allo spettro d'insieme sia legittima, va fatta un'analisi analoga a quella già vista nell'Esempio 1.5.1. Solo allora si può stabilire l'ergodicità e dedurre pertanto lo spettro di potenza di un processo da quello di una delle sue realizzazioni.

Lo studio dello spettro di potenza può essere utilizzato per approfondire la conoscenza di un processo stocastico, analizzato finora nella sua funzione di auto-correlazione. A questo scopo, senza alcuna pretesa di precisione grafica, nella Figura 1.6.2, si riportano gli spettri di potenza relativi alle quattro funzioni di auto-correlazione della Figura 1.6.1. A riguardo, si può osservare che, laddove sono presenti componenti oscillatorie, lo spettro di potenza ne rileva le frequenze, quali puri impulsi (come nel caso d) o quali campane centrate sulle frequenze che interessano (come nel caso c, dove alle componenti oscillatorie sono sovrapposti fenomeni casuali).

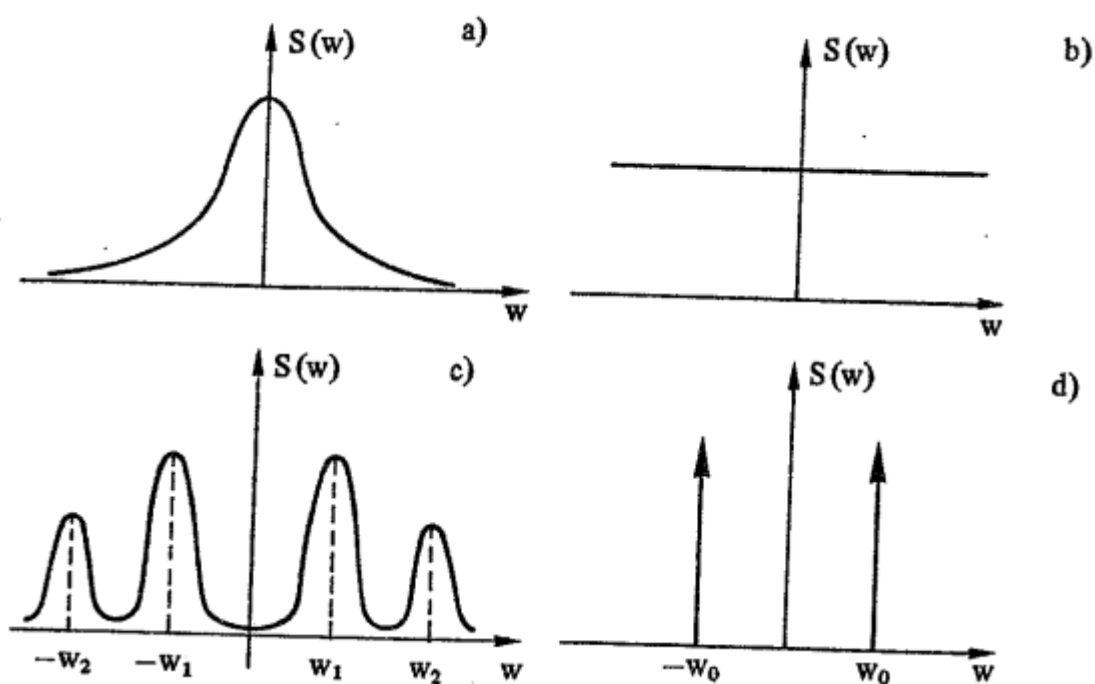


Fig. 1.6.2 – Diversi tipi di spettri di potenza  
(corrispondenti agli auto-correlogrammi di figura 1.6.1)

## 1.7 Esempi di processi stocastici

Alcuni fra i più classici esempi di processi stocastici sono scelti più con il criterio di chiarire i concetti fondamentali che per la loro rilevanza nell'ambito dei processi stocastici. Infatti sono qui omessi processi stocastici di importanza fondamentale, come quelli di Markov, Wiener-Lévy, ecc., per i quali si rimanda alla letteratura specializzata.

### 1.7.1. Processo di Poisson

La rappresentazione grafica di alcune realizzazioni del processo è data in Figura 1.7.1 ed il processo stocastico è così definito:

$$X(0) = 0 \quad \text{e} \quad X(t_1) - X(t_2) = \text{numero di arrivi nell'intervallo}(t_2, t_1)$$

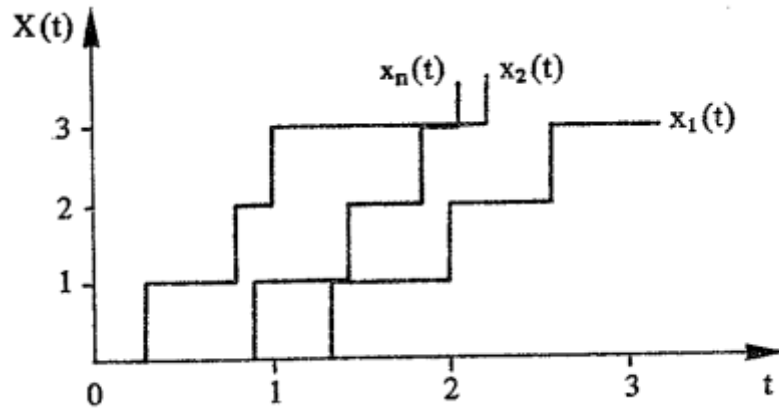


Fig. 1.7.1 – Rappresentazione grafica del processo di Poisson

Gli arrivi degli eventi che seguono la distribuzione di Poisson sono rappresentati da salti unitari negli istanti dell'arrivo. Ad esempio, se degli elettroni sono emessi in modo casuale ed un contatore li registra, l'indicazione del contatore  $t$  in è  $X(t)$ .

Dati  $t_a$  e  $t_b < t_a$ , la variabile casuale  $X(t_a) - X(t_b)$  è distribuita in modo Poissoniano, con parametro  $\lambda(t_a - t_b)$ , per cui:

$$M(X(t_a) - X(t_b)) = \lambda(t_a - t_b) \quad (7.1)$$

Se  $t_a > t_b > t_c > t_d$ , le variabili casuali  $X(t_a) - X(t_b)$  e  $X(t_c) - X(t_d)$  sono indipendenti, cosicchè:

$$M((X(t_a) - X(t_b))(X(t_c) - X(t_d))) = \lambda^2(t_a - t_b)(t_c - t_d)$$

Se invece gli intervalli  $(a, b)$  e  $(c, d)$  sono parzialmente sovrapposti, con un banalissimo calcolo, si ottiene:

$$M((X(t_a) - X(t_b))(X(t_c) - X(t_d))) = \lambda^2(t_a - t_b)(t_c - t_d) + \lambda(t_b - t_c) \quad (7.2)$$

dove  $t_b - t_c$  è la lunghezza della parte sovrapposta. Si poi si pone:  $t_a = t$  e  $t_b = 0$ , la (7.1) diventa:

$$M(X(t)) = \lambda t \quad (7.3)$$

e se invece si pone:  $t_a = t_1$ ,  $t_c = t_2$ ,  $t_b = t_d = 0$ , la (7.2) diventa:

$$\begin{aligned} R_{XX}(t_1, t_2) = M(X(t_1)X(t_2)) &= \lambda t_2 + \lambda^2 t_1 t_2 & t_1 \geq t_2 \\ &= \lambda t_1 + \lambda^2 t_1 t_2 & t_1 \leq t_2 \end{aligned} \quad (7.4)$$

Le (7.3), (7.4) rappresentano la media e l'auto-correlazione del processo di Poisson il quale, evidentemente, non è un processo stazionario, neppure in senso lato.

### 1.7.2. Segnali telegrafici semicasuali

Un processo stocastico  $X(t)$  il cui andamento è come quello della Figura 1.7.2 ha valori:

- $X(t) = 1$ , se il numero totale di arrivi, nell'intervallo  $(0, t)$ , è pari;
- $X(t) = -1$ , se questo numero è dispari,

con la probabilità di avere  $k$  arrivi, nell'intervallo  $(0, t)$ , data da:

$$p_k = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

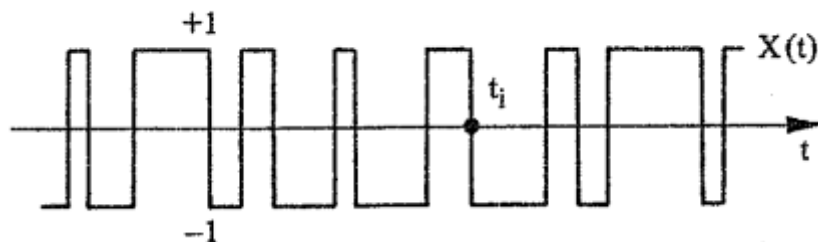


Fig. 1.7.2 – Processo stocastico dei segnali telegrafici semicasuali

Infatti la probabilità di avere un numero pari di arrivi, nell'intervallo  $(0, t)$ , è:

$$P(X(t) = 1) = p(0) + p(2) + \dots = e^{-\lambda t} \left( 1 + \frac{(\lambda t)^2}{2!} + \dots \right) = e^{-\lambda t} \cosh \lambda t$$

ed analogamente la probabilità di avere un numero dispari di arrivi, nell'intervallo  $(0, t)$ , è:

$$P(X(t) = -1) = p(1) + p(3) + \dots = e^{-\lambda t} \left( 1 + \frac{(\lambda t)^3}{3!} + \dots \right) = e^{-\lambda t} \sinh \lambda t$$

La media d'insieme del processo si ottiene dalla 2.1:

$$M(X(t)) = (+1)e^{-\lambda t} \cosh \lambda t + (-1)e^{-\lambda t} \sinh \lambda t = e^{-2\lambda t}.$$

Il calcolo dell'auto-correlazione richiede la distribuzione bidimensionale di  $X(t_1)$  e  $X(t_2)$  che si ricava una volta note le distribuzioni condizionate. Posto  $\tau = t_2 - t_1$ , se  $X(t_1) = 1$ , si ha  $X(t_2) = 1$ , solo se accade un numero pari di arrivi nell'intervallo  $(t_1, t_2)$  da cui:

$$P((X(t_2) = 1) / X(t_1) = 1) = e^{-\lambda \tau} \cosh \lambda \tau \tag{7.5}$$

Moltiplicando la (7.5) per  $P(X(t_1) = 1)$ , si ha la probabilità congiunta:

$$P((X(t_2) = 1) \cap (X(t_1) = 1)) = e^{-\lambda\tau} \cosh \lambda\tau \cdot e^{-\lambda t_1} \cosh \lambda t_1 \quad (7.6)$$

Allo stesso modo, si ottengono le probabilità dei rimanenti tre possibili stati del processo bidimensionale:

$$\begin{aligned} P((X(t_2) = -1) \cap (X(t_1) = -1)) &= e^{-\lambda\tau} \cosh \lambda\tau \cdot e^{-\lambda t_1} \sinh \lambda t_1 \\ P((X(t_2) = 1) \cap (X(t_1) = -1)) &= e^{-\lambda\tau} \sinh \lambda\tau \cdot e^{-\lambda t_1} \sinh \lambda t_1 \\ P((X(t_2) = -1) \cap (X(t_1) = 1)) &= e^{-\lambda\tau} \sinh \lambda\tau \cdot e^{-\lambda t_1} \cosh \lambda t_1 \end{aligned} \quad (7.7)$$

Infine moltiplicando i valori  $1$  o  $-1$ , possibili valori del prodotto  $X(t_1) \cdot X(t_2)$ , per le rispettive probabilità, e sommando, si ottiene l'auto-correlazione del processo stocastico:

$$R_{XX}(t_1, t_2) = M(X(t_1)X(t_2)) = e^{-2\lambda|t_1 - t_2|} = e^{-2\lambda|\tau|} \quad (7.8)$$

### 1.7.3 Segnali telegrafici casuali

Un nuovo processo stocastico:  $Y(t) = aX(t)$ , è ricavato dal precedente, dove  $a$  è una variabile casuale, con media nulla:  $M(a) = 0$ , e varianza unitaria  $\sigma_a^2 = 1$ , i cui valori sono  $+1$  e  $-1$ , con uguale probabilità:

$$a \begin{cases} 1 & -1 \\ 1/2 & 1/2 \end{cases}$$

Supponendo inoltre che  $a$  sia indipendente da  $X(t)$  per qualunque  $t$ , il nuovo processo  $Y(t)$  è uguale ad  $X(t)$  oppure ad  $-X(t)$ , cosicché:

$$\begin{aligned} M(Y(t)) &= M(a)M(X(t)) = 0 \\ R_{YY} &= M[Y(t_1)Y(t_2)] = M[a^2 X(t_1)X(t_2)] = e^{-2\lambda(t_2 - t_1)} = e^{-2\lambda|\tau|} \end{aligned}$$

Il processo è stazionario anche nella media, potendosi dimostrare che è anche ergodico nella media, ed il suo spettro di potenza di è dato da:

$$S(w) = \int_{-\infty}^{+\infty} e^{-2\lambda|\tau|} e^{-jw\tau} d\tau = \frac{4\lambda}{4\lambda^2 + w^2}$$

Le Figure 1.6.1.a e 1.6.2.a mostrano le rappresentazioni grafiche della funzione di auto-correlazione e dello spettro di potenza i cui valori massimi, delle rispettive ordinate, sono di  $1$  e  $1/\lambda$ . Il processo dei segnali telegrafici casuali si riferisce a particolari processi di decadimento, regolati dalla legge di Poisson.

#### 1.7.4. Fenomeno ondulatorio semplice

Date  $n$  variabili casuali  $a_i$ , con le seguenti statistiche:

$$M(a_i) = 0 \qquad \sigma^2(a_i) = \sigma_i^2 \qquad M(a_i a_k) = 0$$

il processo stocastico, caratteristico di un fenomeno ondulatorio semplice, ha forma:

$$X(t) = \sum_{i=1}^n a_i e^{jw_i t}$$

e le sue statistiche di primo e secondo ordine, nonché il suo spettro di potenza, sono rispettivamente:

$$M(X(t)) = \sum_{i=1}^n M(a_i) e^{jw_i t} = 0$$

$$R_{XX}(t_1, t_2) = M(X(t_1) \cdot X^*(t_2)) = \sum_i \sum_k M(a_i a_k) e^{jw_i t_1} e^{-jw_k t_2} =$$

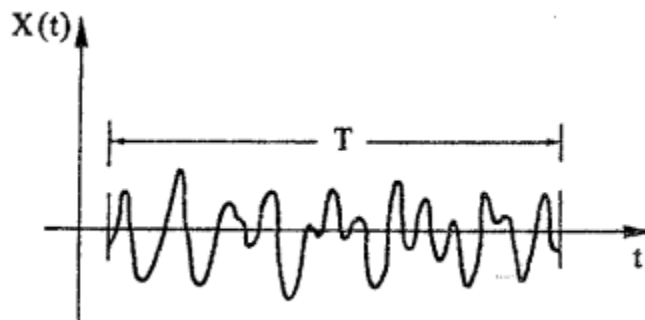
$$= \begin{cases} \sum_i \sigma_i^2 e^{jw_i(t_1-t_2)} = \sum_i \sigma_i^2 e^{jw_i \tau} & \text{se } i = k \\ 0 & \text{se } i \neq k \end{cases}$$

$$S(w) = \int_{-\infty}^{+\infty} \left( \sum_i \sigma_i^2 e^{jw_i \tau} \right) e^{jw_i \tau} d\tau = 2\pi \sum_i \sigma_i^2 \delta(w - w_i)$$

Va notato che l'auto-correlazione  $R_{XX}$  non tende a zero (al crescere di  $\tau$ ), ma si riproduce periodicamente, rivelando che  $X(t)$  non è un vero processo stocastico. Inoltre in  $R_{XX}$ , le fasi delle varie cosinusoidi sono tutte nulle, come deve essere per avere la simmetria intorno a  $\tau = 0$ , mentre le sue ampiezze sono tutte proporzionali alle varianze delle sinusoidi, ovvero alle rispettive energie.

Lo spettro di potenza è invece costituito da una serie di impulsi, nei punti dell'asse  $\omega$ , corrispondenti alle varie  $\omega_i$ , che sono proporzionali alle rispettive  $\sigma_i^2$ .

La Fig. 1.7.3 riporta l'andamento di un particolare  $X(t)$ , costituito da una sinusoide più sette armoniche, della sua funzione di auto-correlazione e del suo spettro di potenza.



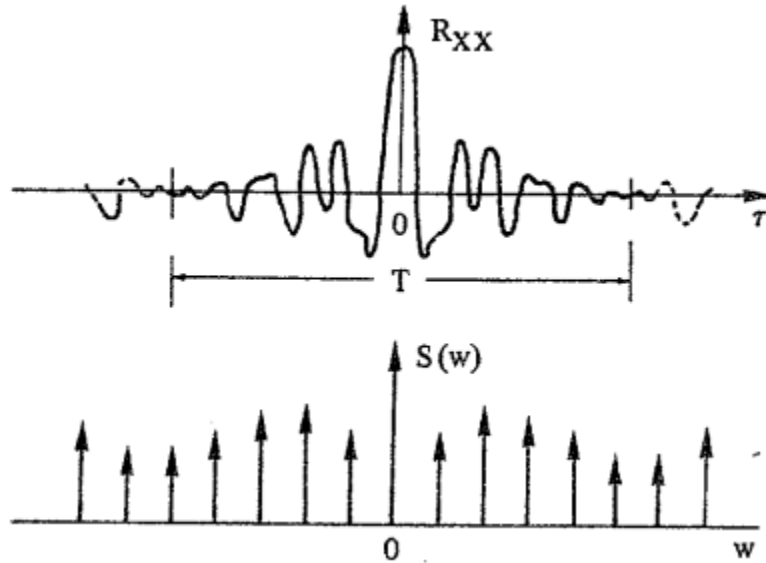


Fig. 1.7.3 – Una realizzazione, l'auto-correlazione e lo spettro di potenza di un processo stocastico puramente oscillatorio

### 1.7.5 Fenomeno ondulatorio stocastico

Il processo stocastico:  $X(t) = a(t)\text{sen}(\omega_0 t + \varphi)$ , è un processo ondulatorio stocastico, nel quale  $\omega = \text{cost}$ , mentre  $a(t)$  è un processo ergodico del tipo già visto nel paragrafo 1.7.3., con le seguenti statistiche:

$$M(a(t)) = 0 \quad R_{aa}(\tau) = a^2 e^{-2\lambda|\tau|}$$

e  $\varphi$  è una variabile casuale, indipendente da  $a(t)$ , con distribuzione omogenea, sull'intervallo  $(0, 2\pi)$ :

$$f(\varphi) = \begin{cases} \frac{1}{2\pi} & 0 \leq \varphi \leq 2\pi \\ 0 & \text{altrove} \end{cases}$$

La media e l'auto-correlazione del processo stocastico sono:

$$M(X(t)) = M(a(t))M(\text{sen}(w_0 t + \varphi)) = 0 \quad (7.9)$$

$$\begin{aligned} R_{XX}(\tau) &= M((a(t+\tau)a(t))(\text{sen}(w_0 t + w_0 \tau + \varphi)\text{sen}(w_0 t + \varphi))) = R_{aa}(\tau) \int_0^{2\pi} \text{sen}(\bullet)\text{sen}(\bullet) \frac{1}{2\pi} d\varphi = \\ &= a^2 e^{-2\lambda|\tau|} \frac{1}{2\pi} \int_0^{2\pi} \text{sen}(w_0 \tau + \varphi)\text{sen} \varphi \quad d\varphi = a^2 e^{-2\lambda|\tau|} \frac{1}{2\pi} \pi \cos w_0 \tau = \frac{a^2}{2} e^{-2\lambda|\tau|} \cos w_0 \tau \end{aligned}$$

L'autocorrelazione trovata è del tipo esponenziale-coseno e, come già precedentemente menzionato, è tipica di quei processi in cui confluiscono fenomeni accidentali ed oscillatori armonici, ad esempio, come il comportamento di un oscillatore meccanico, sollecitato da impulsi casuali, dovuti a moti browniani.

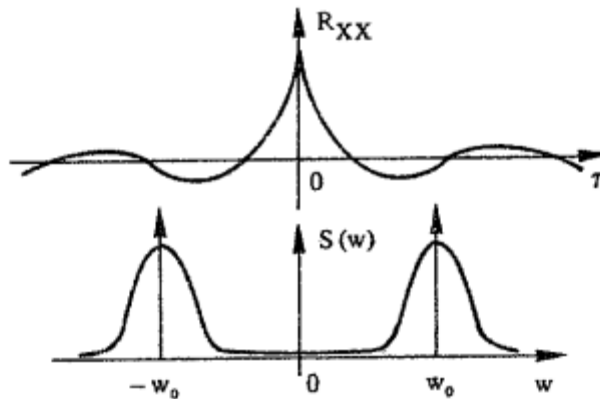


Fig. 1.7.4 – Auto-correlazione e spettro di potenza di un processo ondulatorio stocastico

Lo spettro di potenza dello stesso processo stocastico è dato da:

$$\begin{aligned}
 S(w) &= \int_{-\infty}^{+\infty} \frac{a^2}{2} e^{-2\lambda|\tau|} \cos w_0 \tau e^{-jw\tau} d\tau = 2 \frac{a^2}{2} \int_{-\infty}^{+\infty} e^{-2\lambda|\tau|} \cos w_0 \tau \cos w \tau d\tau = \quad (7.10) \\
 &= \frac{a^2}{2} \int_{-\infty}^{+\infty} e^{-2\lambda|\tau|} [\cos(w + w_0)\tau + \cos(w - w_0)\tau] d\tau = \frac{a^2}{2} \left\{ \frac{2\lambda}{4\lambda^2 + (w + w_0)^2} + \frac{2\lambda}{4\lambda^2 + (w - w_0)^2} \right\}
 \end{aligned}$$

La figura Fig. 1.7.4 rappresenta graficamente l'auto-correlazione e lo spettro di potenza del processo stocastico. Nello spettro di potenza, compaiono due campane i cui massimi sono rispettivamente in  $-\omega_0$  e  $\omega_0$ , dove la maggiore o minore altezza delle campane dipende dal fattore di smorzamento  $\lambda$ .

## PARTE II – ESTENSIONI DELLA DISTRIBUZIONE DI POISSON

### 2.1 Distribuzione esponenziale

Nella distribuzione di Poisson è possibile interpretare il parametro  $\lambda$  come numero medio di eventi nell'unità di tempo (o di spazio). Volendo segnalare che gli intervalli di tempo considerati possono essere costituiti da un numero qualsiasi di intervalli unitari, si può riscrivere la funzione delle probabilità elementari:

$$P_k = \frac{\lambda^k}{k!} e^{-\lambda} \quad \Rightarrow \quad p_x(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \quad \text{per ogni } x = 0, 1, 2, \dots \quad (1.1)$$

evidenziando il legame fra la distribuzione di Poisson ed il processo di Poisson  $X(t)$ , dipendente dal tempo, esaminato nel punto 1.7.1, come esempio di processo stocastico. La forma di  $X(t)$  presenta valori, in ogni momento  $t$ , dati dal numero casuale di eventi, presenti a partire dal momento  $t = 0$ , posto che la distribuzione degli eventi stessi sia di tipo Poissoniano. Ognuna delle spezzate rappresenta una particolare realizzazione del processo che procede con salti di una unità in corrispondenza dei tempi in cui si

presentano gli eventi. A riguardo, per generare un processo di Poisson, il meccanismo fisico, origine degli eventi, deve soddisfare alle note condizioni di stazionarietà, non molteplicità ed indipendenza.

In ogni momento  $t = t_0$ , il valore di  $X(t)$  è una variabile casuale distribuita in modo Poissoniano, secondo la (1.1), con parametro  $\lambda t_0$  e, in generale, il valore  $\lambda$  si chiama la velocità media di arrivo degli eventi nel processo. Dopodiché la distribuzione della variabile casuale  $T(t)$  spiega il tempo intercorso fra zero ed il primo arrivo, cioè la distribuzione da associare all'insieme dei valori  $t$ , ottenuti intersecando il processo di Poisson con la retta:  $X(t) = 1$ .

La probabilità che  $T$  superi un prefissato istante  $t$  è uguale a quella che nessun evento si presenti nell'intervallo  $(0, t)$  che è data da:  $1 - F_T(t)$  (dove:  $F_T(t) = P(T < t)$ ). Questa probabilità è anche esprimibile come  $p_x(0)$ , cioè come la probabilità che una variabile casuale  $X$ , distribuita in modo Poissoniano con parametro  $\lambda t$ , assuma valore nullo:

$$1 - F_T(t) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t} \quad \text{per} \quad t \geq 0$$

$$F_T(t) = 1 - e^{-\lambda t} \quad \text{e} \quad f_T(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t} \quad (1.2)$$

La (1.2) definisce la cosiddetta distribuzione esponenziale che rappresenta il tempo fino al verificarsi del primo evento di tipo Poissoniano. Tuttavia a seguito delle proprietà di stazionarietà ed indipendenza del processo  $e^{-\lambda t}$  è anche la probabilità che non si verifichino eventi in nessun intervallo di durata  $t$  (che inizi o meno al tempo zero).

Di conseguenza, se si pone come origine dei tempi il momento in cui si verifica l' $n$ -esimo evento,  $e^{-\lambda t}$  rappresenta la probabilità che il tempo intercorso fra gli eventi  $n$ -esimo e  $n + 1$ -esimo sia superiore a  $t$ , cosicché i tempi fra arrivi successivi, in un processo di Poisson, sono indipendenti e distribuiti in modo esponenziale.

Ponendo  $u = \lambda t$ , la media della distribuzione esponenziale è:

$$M(T) = \int_0^{\infty} t \lambda e^{-\lambda t} dt = \frac{1}{\lambda} \int_0^{\infty} u e^{-u} du = \frac{1}{\lambda} [e^{-u}(-u - 1)]_0^{\infty} = \frac{1}{\lambda} \quad (1.3)$$

dove il valore  $1/\lambda$  rappresenta il tempo medio fra arrivi successivi. Analogamente la varianza della stessa distribuzione è:

$$\sigma^2(T) = 1/\lambda^2 \quad (1.4)$$

La rappresentazione grafica della distribuzione esponenziale è data nella Fig. 2.1.1, in funzione di  $\lambda t$ , rapporto fra  $t$  ed il tempo medio fra arrivi successivi.

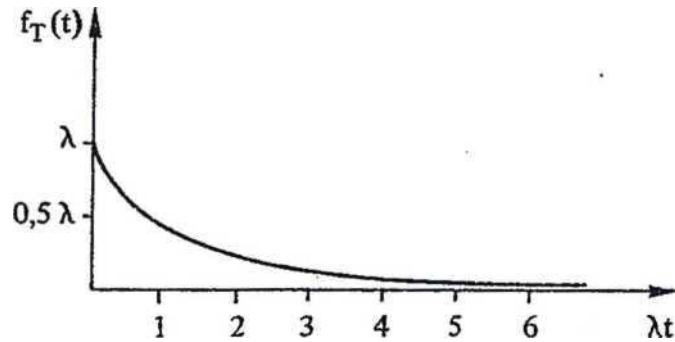


Fig. 2.1.1 – Densità di probabilità di tipo esponenziale

## 2.2 Assenza di memoria del processo di Poisson

Il processo di Poisson è senza memoria, nel senso che il suo futuro comportamento è indipendente da quello presente o da quello passato, come mostrato dalla distribuzione di  $T(t)$  condizionata a  $T > t_0$ , cioè dalla distribuzione del tempo fra arrivi successivi, posto che nessun evento si verifichi prima di  $t_0$ :

$$F_{T/(T>t_0)}(t) = P((T \leq t)/(T > t_0)) = \frac{P((T \leq t) \cap (T > t_0))}{P(T > t_0)}$$

Per  $t < t_0$ , il numeratore è nullo, mentre per  $t > t_0$ , è semplicemente:  $P(t_0 < T < t)$ , da cui:

$$F_{T/(T>t_0)}(t) = \frac{F_T(t) - F_T(t_0)}{1 - F_T(t_0)} = \frac{1 - e^{-\lambda t} - (1 - e^{-\lambda t_0})}{e^{-\lambda t_0}} = 1 - e^{-\lambda(t-t_0)} \quad t \geq t_0 \quad (2.1)$$

$$f_{T/(T>t_0)}(t) = \lambda e^{-\lambda(t-t_0)} \quad t \geq t_0 \quad (2.2)$$

e, se si misura il tempo a partire da  $t_0$ , ovvero da  $\tau = t - t_0$ , la (2.2) diventa:

$$f_{T/(T>t_0)}(\tau) = \lambda e^{-\lambda \tau} \quad \text{per ogni } \tau \geq 0$$

Questo significa che, anche se non si sono verificati eventi fino a  $t_0$ , le previsioni, fatte sul tempo, prima del primo arrivo, a partire da  $t_0$ , non cambiano, essendo sempre regolate da una distribuzione esponenziale con parametro  $\lambda$  (cosicché il processo di Poisson è invariante rispetto alla scelta dell'origine).

## 2.3 Distribuzione Gamma di Erlang

Se si interseca il processo di Poisson con la retta:  $X(t) = k$ , si ottiene la serie dei valori argomentali della valori argomentali dei tempi che intercorrono da zero fino al  $k$ -esimo arrivo. Sia  $X_k$  questa variabile casuale, si può considerare  $X_k$  nella forma  $X_k = T_1 + T_2 + \dots + T_k$ , dove  $T_i$  è la variabile casuale dei

tempi di arrivo, fra gli eventi  $(i-1)$ -esimo ed  $i$ -esimo, e tutte le variabili casuali hanno la stessa distribuzione esponenziale. Allora la variabile casuale  $X_k$  è la somma di più variabili casuali di distribuzione nota la cui distribuzione si trova con una ripetuta applicazione dell'integrale di convoluzione:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx$$

Infatti supposto:  $X_2 = T_1 + T_2$ , con

$$f_{T_1}(v) = \lambda e^{-\lambda v} \quad \text{e} \quad f_{T_2}(w) = \lambda e^{-\lambda w}$$

si ha:

$$f_{x_2}(x) = \int_{-\infty}^{+\infty} f_{T_1}(v)f_{T_2}(x-v)dv = \int_0^x \lambda e^{-\lambda v} \lambda e^{-\lambda(x-v)} dv = \lambda^2 \cdot e^{-\lambda x} \int_0^x dv = \lambda^2 x e^{-\lambda x} = \frac{\lambda(\lambda x)e^{-\lambda x}}{(2-1)!}$$

e nello stesso modo, procedendo per  $X_3, \dots, X_k$ , si ricava la distribuzione Gamma di Erlang:

$$f_{xk} = \frac{\lambda(\lambda x)^{k-1} e^{-\lambda x}}{(k-1)!} \quad \text{per ogni } x \geq 0 \quad (3.1)$$

Dalla definizione di  $X_k$ , si ricavano la sua media, la sua varianza ed il suo indice di asimmetria:

$$M(X_k) = k/\lambda \quad \sigma^2(X_k) = k/\lambda^2 \quad \text{e} \quad \gamma = 2/\sqrt{k}$$

L'andamento di questa distribuzione, rappresentato in Fig. 2.3.1, dipende da due parametri:  $\lambda$  che si può considerare un parametro di scala e  $k$  che è invece un parametro di forma. In realtà, la distribuzione Gamma è definita più ampiamente di quanto implichino la (3.1), con la sola restrizione:  $\lambda$  e  $k > 0$ :

$$f_X(x) = \frac{\lambda(\lambda x)^{k-1} e^{-\lambda x}}{\Gamma(k)} \quad \text{per ogni } x \geq 0 \quad (3.2)$$

dove la funzione  $\Gamma(k)$  vale  $(k-1)!$ , se  $k$  è intero, ed è così definita, in generale:

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

funzione tabulata, come *Gamma incompleta*:

$$\Gamma(k, x) = \int_0^x e^{-u} u^{k-1} du \quad \text{ed anche: } \Gamma(k, \lambda x) = \Gamma(k)F_{X_k}(x) = \Gamma(k) \int_0^x f_{X_k}(x) dx$$

dove l'ultima espressione può essere usata per trovare la funzione di distribuzione:  $F_{X_k}(x)$ .

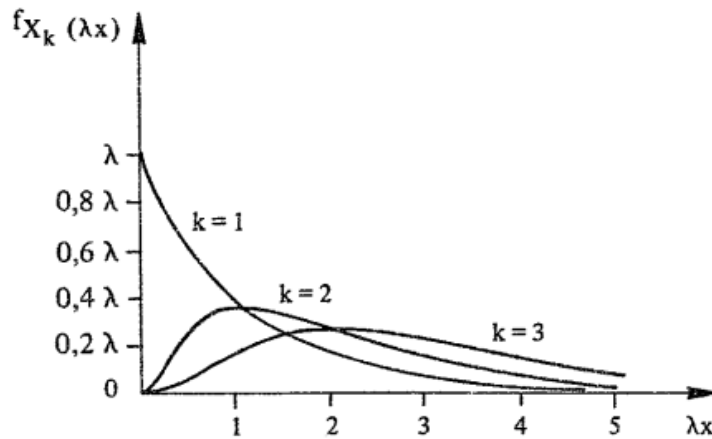


Fig. 2.3.1 – Densità di probabilità della variabile casuale *Gamma* di Erlang

Le espressioni della media, della varianza e dell'indice di asimmetria valgono anche nel caso di  $k$  non intero. La distribuzione Gamma di Erlang nel suo significato originale si presta a descrivere il tempo di arrivo di un prefissato numero di veicoli od il momento del cedimento di una struttura, progettata per sopportare un determinato numero di sovraccarichi, prima di crollare, posto che gli arrivi dei veicoli o la posa in opera dei sovraccarichi seguano le condizioni di stazionarietà, non molteplicità ed indipendenza, richieste per gli eventi poissoniani. Tuttavia la distribuzione Gamma di Erlang applicata estensivamente anche in svariati altri campi dell'ingegneria, a fenomeni come le portate massime di corsi d'acqua, l'entità delle precipitazioni mensili e la resistenza al carico di strutture in cemento armato. Questo uso Gamma è fatto indipendentemente dalla sua derivazione logica, perché si è rivelata un buon modello matematico di questi fenomeni concreti. In questo contesto, è frequentemente usata la cosiddetta *Gamma traslata* la cui densità di probabilità è:

$$f_x(x) = \frac{\lambda(\lambda(x-a))^{k-1} e^{-\lambda(x-a)}}{\Gamma(k)} \quad (3.3)$$

la cui derivazione logica proviene dalla constatazione che in molti casi, il valore della variabile casuale  $x$  non può essere inferiore ad un certo valore  $a$ . La Gamma traslata dipende non più da due, ma da tre parametri  $\lambda$ ,  $k$  ed  $a$ , ed anzitutto, quando si pone il problema di ricercare, se questa distribuzione può servire come modello matematico, per un certo fenomeno, bisogna determinare, in base ai dati osservati fino a quel momento, quali sono i valori dei parametri da adottare nel caso specifico. In generale, si adotta il metodo dei *momenti*, nonostante questo metodo di stima non sia molto soddisfacente dal punto di vista delle proprietà degli stimatori. Pertanto in base ai dati, si calcolano:  $\bar{x}$ ,  $\sigma^2$  e  $\gamma$ , risolvendo uno dei due sistemi (il primo valido per stimare i parametri della (3.2) ed il secondo per quelli della (3.3)):

$$\begin{cases} \bar{x} = k / \lambda \\ \sigma^2 = k / \lambda^2 \end{cases} \quad \begin{cases} \bar{x} = a + k / \lambda \\ \sigma^2 = k / \lambda^2 \\ \gamma = 2 / \sqrt{k} \end{cases} \quad (3.4)$$

Ovviamente è possibile ricavare i parametri delle (3.4) applicando il metodo di massima verosimiglianza che tuttavia richiede calcoli notevolmente più onerosi. Infatti eventuali controlli di buona interpolazione

richiedono l'uso di tabulazioni della distribuzione Gamma incompleta non sempre accessibili. In alternativa, si ricorre alle tabulazioni della variabile casuale  $\chi^2$ , di molto più facile reperimento, sfruttando il fatto che ogni curva di Figura 2.3.1 coincide con la funzione densità di probabilità  $f(\chi^2)$ , con  $\nu = 2k$  gradi di libertà, purché si esegua la trasformazione:  $\chi^2 = 2\lambda x$ .

### Esempio 2.3.1

Nell'organizzazione delle varie fasi di rilevanti lavori edili e del numero dei mezzi meccanici necessari per scavi, trasporti e riporti, è opportuno lo studio del tempo di carico di un autocarro da parte di una pala meccanica. Riuscire a stabilire un modello matematico per variabile casuale *tempo di carico*, può, essere utile, per decidere quanti autocarri sono necessari all'impresa, per ridurre, ad un minimo, i tempi di attesa degli autocarri e della pala.

A tal fine, i dati di un campione casuale di 20 tempi di carico sono sistemati in ordine crescente, per determinare i parametri della distribuzione Gamma di Erlang, utilizzata per rappresentare il fenomeno e controllare la bontà dell'adattamento.

1.23 1.35 1.41 1.43 1.47 1.48 1.53 1.55 1.58 1.59  
1.62 1.63 1.67 1.71 1.75 1.79 1.84 1.95 2.12 2.33

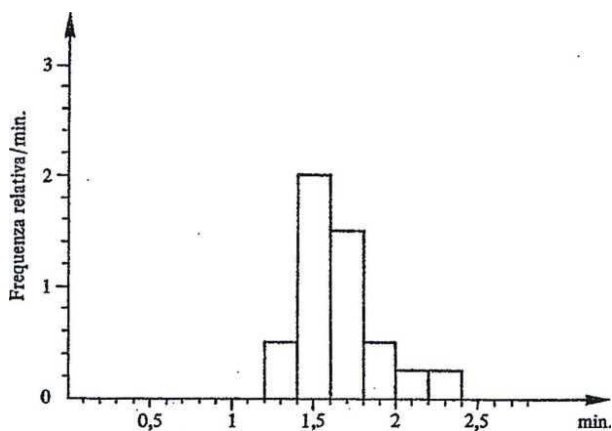


Fig. 2.3.2 – Istogramma relativo a 20 tempi di carico

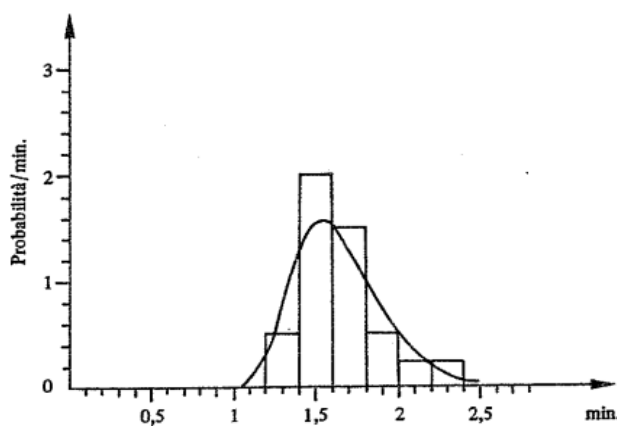


Fig. 2.3.3 – Distribuzione Gamma traslata adattata ai dati sperimentali

I valori del campione, raccolti sotto forma di istogramma (come mostra la Figura 2.3.2) rivelano una discreta asimmetria, con la prevalenza dei tempi elevati, per cui la distribuzione Gamma di Erlang (con asimmetria positiva) potrebbe costituire un modello matematico adeguato e così, in base ai dati campionari, si calcola:  $\bar{t} = 1.6515$ ,  $\sigma = 0.2627$  e  $\gamma = 0.8566$ .

La media e la varianza campionarie permettono di stimare i parametri  $\lambda$  e  $k$  della distribuzione, in quanto:  $k/\lambda = \bar{t} = 1.6515$  e  $\sqrt{k}/\lambda = \sigma = 0.2627$ , cosicché si ha, risolvendo il sistema, si ha:  $k = 39.52$  e  $\lambda = 23.93$ , da cui si ricava il coefficiente di asimmetria:  $\gamma = 2/\sqrt{k} = 0.32$ , valore molto diverso da quello stimato in base ai dati.

Allora è opportuno introdurre un altro tipo di distribuzione Gamma di Erlang che tenga conto di un certo tempo minimo, necessario per caricare l'automezzo. Pertanto si deve usare una distribuzione di tipo (3.3) nella quale anche il parametro  $a$  deve essere stimato, in base ai dati campionari, cosicché ora le equazioni sono tre:

$$\bar{t} = a + k/\lambda = 1.6515 \quad \sigma = \sqrt{k}/\lambda = 0.2627 \quad \text{e} \quad \gamma = 2/\sqrt{k} = 0.8566$$

e da queste si ottengono le stime dei parametri:  $k = 5.45$ ,  $\lambda = 8.89$  e  $a = 1.04$ .

In Figura 2.3.3, è riportato l'istogramma di Figura 2.3.2, con sovrapposta la funzione densità di probabilità Gamma traslata, assunta come modello.

Resta ora da verificare, se la distribuzione ipotizzata è accettabile, come modello matematico del fenomeno, oppure se i dati sperimentali si discostano invece significativamente da questo.

A tal fine, è necessario un controllo sul buon adattamento dell'interpolazione, mediante un test classico, facente uso della distribuzione  $\chi^2$ . I valori delle probabilità riportati nella terza colonna della tabella seguente sono ricavati utilizzando, in luogo della tabulazione della distribuzione Gamma di Erlang quella della equivalente  $\chi^2$ , con  $\nu = 2k = 10.90$  gradi di libertà:

$$\chi^2 = 2\lambda(x - a) = 17.77(x - 1.04)$$

Il confronto fra le frequenze osservate  $f_0$  e quelle teoriche  $f_t$  porta ad un valore.

$$\chi^2_{\text{sperim.}} = \sum (f_0 - f_t)^2 / f_t = 2.899 \quad \text{con} \quad \nu = 9 - 3 - 1 = 5$$

Le loro differenze non sono significative a nessuno dei livelli comunemente accettati e così è possibile accettare, come modello matematico per l'esperimento sotto esame, l'equazione:

$$f(t) = \frac{8.89^{5.45}}{\Gamma(5.45)} (t - 1.04)^{4.45} e^{-8.89(t-1.04)} \quad \text{dove} \quad \Gamma(5.45) \cong 50$$

$x$	$\chi^2$	$1 - \chi \int_0^{\chi^2} f(\chi^2) d\chi^2$	$\Delta$	$f_t = 20\Delta$	$f_0$
1.04	0	1			
1.24	3.55	0.9796	0.0204	0.408	1
1.44	7.11	0.7825	0.1971	3.942	3
1.64	10.66	0.4636	0.3189	6.378	8
1.84	14.22	0.2156	0.2480	4.960	4
2.04	17.77	0.0844	0.1312	2.624	2
2.24	21.32	0.0294	0.0550	1.100	1
2.44	24.88	0.0092	0.0202	0.404	1
2.64	28.43	0.0027	0.0065	0.130	0
> 2.64	> 28.43	0	0.0027	0.050	0

## APPENDICE A – CONTROLLO STATISTICO DELLA QUALITÀ <sup>2</sup>

### 1. Il miglioramento della qualità

Questa appendice presenta le modalità di impiego di diversi metodi statistici, per la soluzione di problemi finalizzati alla valutazione della qualità ed al miglioramento della qualità di prodotti (come beni, servizi, processi, ecc.).

#### 1.1 Significato dei termini “qualità” e “miglioramento della qualità”

Il termine “qualità” può essere definito in relazione a una o più caratteristiche che un prodotto deve avere. Essa costituisce uno dei fattori più rilevanti in un processo di decisione e, di conseguenza, comprendere e migliorare la qualità è un fattore decisivo per la crescita, il miglioramento ed il successo.

##### 1.1.1 Le componenti della qualità

La qualità può essere valutata secondo diversi aspetti:

- prestazione e funzionalità;**
- affidabilità;**
- durata;**
- manutenibilità;**
- aspetti formali;**
- livello di qualità percepito;**
- conformità alle normative.**

La definizione tradizionale del termine qualità si basa sul presupposto che un prodotto deve soddisfare le richieste di coloro che lo utilizzano:

la **qualità** di un prodotto garantisce il suo essere appropriato all'uso.

Due sono gli aspetti generali dell'essere appropriato per l'uso:

- la qualità di progetto, in base al quale tutti i beni e servizi sono prodotti con vari gradi o livelli di qualità;
- la conformità alle normative, relativa a quanto il prodotto risulta conforme alle specifiche che sono richieste dal progetto.

Questa seconda definizione è associata più agli aspetti della conformità alle normative che a quelli della qualità (intrinseca) di progetto, cosicché si preferisce aggiungere anche una, più recente, definizione del termine qualità:

la **qualità** è inversamente proporzionale rispetto alla variabilità.

---

<sup>2</sup> L'autorevole fonte di questa presentazione è il libro: Controllo statistico della qualità, di Douglas C. Montgomery (McGraw-Hill, Milano, 2000).

Questo implica che, se diminuisce la variabilità nelle caratteristiche di un prodotto, aumenta la qualità del prodotto stesso. Minori riparazioni e minori reclami significano una minore rilavorazione ed una riduzione di spreco di tempo, sforzo e costo. Pertanto la qualità è davvero inversamente proporzionale alla variabilità ed i suoi effetti possono essere comunicati usando un linguaggio che ognuno può comprendere. Questo conduce alla seguente definizione di “miglioramento della qualità”:

**il miglioramento della qualità** di un prodotto è la riduzione della sua variabilità.

Infatti la variabilità eccessiva di un prodotto, come pure nel corso del suo processo produttivo, oppure nelle sue prestazioni funzionali, ecc. conduce spesso ad una perdita, cosicché una definizione alternativa è quella che definisce il miglioramento della qualità come riduzione della perdita.

### 1.1.2 Terminologia tecnica nell'ingegneria della qualità

Ogni prodotto possiede un numero di elementi che congiuntamente descrivono quello che l'utilizzatore od il consumatore ritengono qualità dell'oggetto. Questi elementi sono spesso definiti **caratteristiche di qualità** e possono essere di diversi tipi:

- ❑ **fisiche** (ad esempio, lunghezza, peso, tensione elettrica);
- ❑ **sensoriali** ad esempio, sapore, aspetto, colore);
- ❑ **comportamentali nel tempo** (ad esempio, affidabilità, durata, manutenzione).

Le **tecniche della qualità** sono un insieme di attività, operative, gestionali e tecnologiche, utilizzate per assicurarsi che le caratteristiche di qualità di un prodotto siano di livello nominale o richiesto.

Invece talvolta è difficile fornire prodotti che abbiano caratteristiche di qualità sempre uguali. In ciascun elemento prodotto, od al massimo livello dell'aspettativa. La causa principale di questa **variabilità** dalla cui presenza consegue che non si hanno mai due prodotti identici comprende differenze nei materiali, diversità nelle prestazioni e nell'operatività degli strumenti di produzione e differenze nelle modalità di esecuzione del lavoro. Poiché la variabilità può essere descritta solo in termini statistici, i **metodi statistici** hanno un ruolo centrale negli sforzi per il miglioramento della qualità e, nella loro applicazione all'ingegneria della qualità, è abituale classificare i dati delle caratteristiche di qualità in:

- ❑ **attributi**, di solito di tipo discreto, spesso espressi in termini di conteggio;
- ❑ **variabili**, in genere derivate da misure di tipo continuo, quali lunghezza, peso, tensione elettrica.

Le caratteristiche della qualità sono spesso valutate in relazione alle **specifiche**. Per un prodotto manufatto, le specifiche sono le misure stabilite per alcune caratteristiche tangibili dei componenti o sotto-componenti costituenti il prodotto, come pure i valori desiderati per le caratteristiche di qualità del prodotto finale.

Il valore di misura che corrisponde al valore desiderato per una caratteristica di qualità è definito il **valore nominale** per quella caratteristica. Questi valori nominali possono anche costituire un intervallo di valori che tipicamente si ritiene sufficientemente prossimi al valore teorico di riferimento, da non incidere sulla funzione o sulla prestazione del prodotto, se la caratteristica di qualità rientra bene in questo ambito.

Il maggior valore ammissibile (accettabile) per una caratteristica di qualità è definito **limite di specifica superiore**, *Upper Specification Limit (USL)* e il valore minore ammissibile (accettabile) per una caratteristica di qualità è definito **limite di specifica inferiore**, *Lower Specification Limit (LSL)*, mentre alcune altre caratteristiche di qualità hanno limiti di specifica solo unilaterali.

Di solito, le specifiche ed i limiti di specifica sono stabiliti dai progettisti, durante la progettazione tecnica del prodotto e forniti ai reparti di produzione. E' possibile riferirsi a tutto questo, come ad un approccio alla progettazione di genere **non convenzionale** (*over the wall*).

I problemi relativi alla qualità di un prodotto sono maggiori quando è usato questo tipo di approccio, dove le specifiche sono spesso stabilite senza considerare la variabilità dei materiali, dei processi e delle altre parti del sistema. Uno specifico tipo di difetto è detto **non conformità** e questo inconveniente si palesa in prodotti o componenti **non conformi** che non riescono a raggiungere una o più specifiche.

Un prodotto non conforme non è necessariamente inadatto al suo scopo, mentre è considerato **difettoso**, se ha uno o più **difetti**, costituiti da non conformità abbastanza gravi, da impedire, in modo significativo, l'utilizzazione del prodotto stesso.

La progettazione di sistemi di lavorazione ed assemblaggio si rivela importante nel superamento di problemi connessi all'approccio *over the wall* e molti tecnici ricevono oggi una formazione, in queste aree, come parte integrante della loro preparazione professionale. Infatti la recente enfasi sulla **tecnica convergente** ha indotto ad un approccio di team, con specialisti in processi, tecnici della qualità ed esperti in varie discipline, operanti insieme con il progettista, fin dalle prime fasi della procedura progettuale, al fine di avere un miglioramento della qualità.

## 1.2 Metodi statistici per il miglioramento della qualità

Tre sono gli aspetti fondamentali per il miglioramento della qualità:

- ❑ **il controllo statistico di un processo;**
- ❑ **la programmazione degli esperimenti;**
- ❑ **il campionamento in accettazione.**

La **carta di controllo** (di cui la Figura 1.2.1 mostra un esempio) è una delle principali tecniche adottate per il controllo statistico di un processo o, più brevemente, dell'**SPC** (*Statistical Process Control*), e ha:

- ❑ una linea centrale (**CL**), per indicare il punto dove dovrebbe posizionarsi la caratteristica del processo, senza la presenza di fonti di variabilità anomale;
- ❑ due limiti di controllo, uno superiore (**UCL**) ed uno inferiore (**LCL**), con funzioni di controllo, individuati sulla base di alcune semplici considerazioni statistiche.

La carta di controllo è un'utile **tecnica per il monitoraggio del processo** ed un valido strumento per ridurre la variabilità del processo. Infatti in presenza di fonti di variabilità anomale, può accadere di avere alcuni punti al di fuori dei limiti di controllo.

Allora la carta di controllo fornisce un'indicazione utile sulla necessità di effettuare indagini per rimuovere e correggere gli eventuali errori intervenuti.

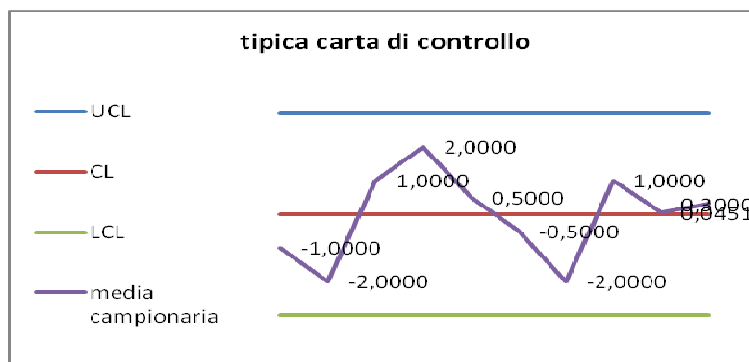


Fig. 1.2.1 – Un esempio di carta di controllo

Un **esperimento programmato** è invece estremamente utile per scoprire quali sono state le variabili che influenzano maggiormente la realizzazione del livello qualitativo osservato e per sapere, in che misura, esso è influenzato dalle variazioni di livello dei fattori studiati. Questo strumento è essenziale, per ridurre tutta la variabilità, a livello della qualità del prodotto, e nel determinare i livelli, assunti delle variabili controllabili, per avere una resa ottimale.

Il tipo più diffuso di esperimento programmato è quello **fattoriale** in cui i livelli dei fattori sono fatti variare, in modo da esaminare l'effetto di tutte le possibili combinazioni dei livelli testati. Fondamentalmente questi esperimenti programmati sono uno strumento per il controllo della qualità di tipo **off-line**, perché sono impiegati nello sviluppo e nella progettazione dell'attività produttiva prima che il processo inizi la sua normale evoluzione, avendo come obiettivo l'individuazione di quei trattamenti che garantiscono la minore variabilità possibile.

Una volta individuate le relazioni funzionali, tra le variabili che regolano la produzione, si possono utilizzare, con grande efficacia, tecniche di monitoraggio e di sorveglianza della produzione. Ad esempio, le carte di controllo possono fornire un'indicazione, di quando un processo ha subito variazioni e della necessità di riportare la produzione sotto controllo. Se gli interventi correttivi sono impiegati, in modo continuativo, si ha un **controllo ingegneristico** o **controllo automatico** o **feedback**.

Il **campionamento in accettazione** è definibile come l'ispezione e la classificazione di un campione di unità, selezionate a caso da un lotto, per la valutazione della qualità del lotto stesso. Le diverse tipologie di campionamento in accettazione sono:

- l'operazione d'**ispezione in uscita**, realizzata immediatamente dopo la produzione e prima della consegna;
- l'operazione d'**ispezione in entrata**, realizzata quando il controllo dei pezzi è effettuato immediatamente prima della loro acquisizione,
- l'operazione d'**ispezione di rettifica**, dove i campioni esaminati possono portare all'accettazione od al rifiuto del lotto e, in quest'ultimo caso, i pezzi esaminati possono essere rilavorati, oppure scartati.

### 1.3 Aspetti gestionali di miglioramento della qualità

Le tecniche statistiche che includono SPC e programmazione degli esperimenti sono, con altri strumenti di risoluzione di problemi, le basi tecniche per il controllo ed il miglioramento della qualità.

La gestione effettiva della qualità comprende la corretta esecuzione di varie attività:

- ❑ **pianificazione della qualità:** un'attività strategica ed altrettanto vitale, per la positiva evoluzione, a lungo termine, di un'organizzazione, in relazione al piano per lo sviluppo del prodotto;
- ❑ **assicurazione della qualità:** insieme delle attività, per garantire che la qualità dei prodotti sia mantenuta ad un livello adeguato e che le questioni, inerenti la qualità, siano risolte opportunamente;
- ❑ **miglioramento della qualità:** insieme delle attività destinate ad assicurare che i prodotti soddisfino le richieste e siano migliorati in modo continuo (poiché la variabilità è la causa principale di cattiva qualità, le tecniche statistiche, incluse SPC e programmazione degli esperimenti, sono i migliori strumenti di controllo e miglioramento).

## 2. Metodi statistici utili nel miglioramento della qualità del processo

Il termine statistica indica un insieme organico di tecniche, impiegabili per prendere decisioni riguardanti un processo, od una popolazione, basandosi sulle informazioni contenute in un campione, prodotto da questo processo o rilevato su questa popolazione. Pertanto i metodi statistici hanno un ruolo importante nel miglioramento della qualità e sono i principali strumenti con cui un prodotto è analizzato, verificato o valutato, in modo da impiegare le informazioni disponibili, al fine di controllare e migliorare il processo di produzione. Inoltre il metodo statistico è un linguaggio mediante il quale gli addetti allo sviluppo, alla produzione, alla gestione e di altri settori comunicano tra loro, con i seguenti obiettivi:

- ❑ definire quali strumenti di statistica descrittiva si possono impiegare, per valutare quantitativamente le variazioni che una caratteristica od un indicatore di qualità presenta tra i possibili valori, considerati in un campione.
- ❑ introdurre distribuzioni di probabilità appropriate e spiegare come queste siano in grado di fornire uno strumento, per la realizzazione del modello, come pure la descrizione delle caratteristiche di qualità di un processo.

### 2.1 La descrizione della variabilità

La descrizione della variabilità si attua essenzialmente attraverso grafici e statistiche.

#### 2.1,1 I grafici "rami e foglie"

Due unità di un bene prodotto, da un processo, non sono mai identiche; infatti qualche variazione è inevitabile e, in questa prospettiva, si può considerare la statistica come la scienza che, dall'analisi dei dati, trae conclusioni, prendendo in considerazione la presenza della variabilità dei dati stessi. Diversi metodi grafici sono utili per sintetizzare e presentare i dati osservati, in un'indagine preliminare. Una di queste tecniche è costituita dai grafici "rami e foglie" (*stem-and-leaf plot*).

A tal fine, si suppone che i dati siano indicati con  $x_1, x_2, \dots, x_n$  e che ciascun numero  $x_i$  sia composto da almeno due cifre. Per ottenere un grafico "rami e foglie", si suddividono le cifre del numero  $x_i$  in due parti: una costituita da una o più cifre iniziali (detta/e il "ramo") e l'altra costituita dalle rimanenti cifre (dette le "foglie").

Ad esempio, si può scindere un valore generico: 76, nella cifra: 7, intesa come il “ramo” e nella cifra: 6, intesa come la “foglia”.

In generale, i valori “rami” sono relativamente pochi rispetto al numero delle osservazioni (di solito un compreso tra cinque e venti). Definito l’insieme dei “rami”, essi sono elencati al margine sinistro, ponendo a fianco tutte le “foglie”, corrispondenti ai valori osservati, nell’ordine in cui sono riscontrate nell’insieme dei dati.

La versione del grafico foglie e rami è talvolta chiamata **grafico rami-foglie ordinato**, poiché le foglie sono disposte per grandezza. Questo modo rende facile trovare i **percentili** dei dati. Infatti in generale, il  $k$ -esimo percentile è un valore tale che, almeno  $k\%$  dei dati sono minori od uguali di questo valore ed almeno  $(1 - k)\%$  dei dati sono maggiori di questo valore. Allora la mediana è il **cinquantesimo percentile**, mentre definiti il **primo quartile** ed il **terzo quartile**, una misura della variabilità è data dalla loro differenza, detta **differenza interquartile** la cui metà, in generale, non è molto dissimile dal median absolute value (mav). Infine benché la rappresentazione, mediante il grafico “rami e foglie”, sia un’ottima modalità per visualizzare la variabilità dei dati, essa non tiene conto dell’**ordine temporale** con cui sono effettuate le osservazioni. Tuttavia spesso il tempo è un fattore importante che contribuisce alla variabilità, in particolare, nei problemi riguardanti il miglioramento della qualità. In questo caso, è altresì possibile rappresentare semplicemente i valori dei dati, rispetto al tempo, in un altro grafico chiamato **grafico della serie temporale** o **carta a scorrimento temporale**.

### 2.1.2 L’istogramma ed il cumulogramma

Un istogramma è una sintesi compatta dei dati simile al grafico rami e foglie che divide il campo di variazione dei dati in intervalli, solitamente chiamati intervalli di classe. Un istogramma che usi un numero eccessivo di classi, come pure troppo limitato, risulta poco informativo per cui, di solito, si usano tra cinque e venti classi, e questo numero dovrebbe aumentare all’aumentare del numero delle osservazioni.

L’istogramma può essere abbastanza sensibile al numero e all’ampiezza delle classi e, nel caso di dati poco numerosi, può addirittura cambiare radicalmente aspetto, cosicché l’istogramma è una tecnica adatta solo per insiemi numerosi di dati. Si noti anche che, nel passare dai dati originari o dal grafico rami e foglie ad un istogramma, si perdono informazioni, perché i valori puntuali delle osservazioni originarie non sono presenti nel grafico. Tuttavia questa perdita di informazioni è piccola, se confrontata con la capacità di sintesi e la facilità di interpretazione dell’istogramma stesso.

Il cumulogramma è ottenuto sommando progressivamente, ad ogni classe, la somma di tutte le sue precedenti, così da produrre un grafico strettamente crescente, da zero (l’insieme vuoto) ad uno (con l’intero universo in esame).

### 2.1.3 Sintesi numerica dei dati

La sintesi numerica dei dati è fornita dalle usuali statistiche campionarie, proprie della statistica descrittiva ad una dimensione. Tra queste sono da segnalare:

- gli indicatori di centro o locazione / posizione, come la media aritmetica e la mediana (di cui la prima ottimale e la seconda robusta), come pure altre medie, geometrica, armonica, ponderata, potata, ecc.;

- ❑ gli indicatori della dispersione, come la varianza (ed il corrispondente scarto quadratico medio, altrimenti detto deviazione standard) e il valore assoluto mediano, o più brevemente mav (di cui ancora la prima ottimale ed il secondo robusto), come pure indici di mutua variabilità (come il delta di Gini e l'entropia di Shannon);
- ❑ gli indicatori di forma, come gli indici di asimmetria e gli indici di curtosi (sul comportamento delle code).

#### 2.1.4 Rappresentazione dei dati mediante “box plot”

Il “box plot” è una rappresentazione grafica che presenta diversi importanti indicatori dei dati osservati, quali tendenza centrale o locazione, dispersione o variabilità, allontanamento dalla simmetria distributiva e identificazione delle osservazioni anomale, distanti dal nucleo centrale dei dati (*outliers*). Il grafico box plot presenta i tre quartili:

- ❑ il valore minimo e il valore massimo, in una scatola rettangolare (disposta orizzontalmente, oppure verticalmente), dove l'ampiezza del rettangolo rappresenta la differenza interquartile con il primo quartile a sinistra (od in basso) ed il terzo quartile a destra (od in alto);
- ❑ una linea intermedia corrispondente al secondo quartile (o mediana).

Due segmenti, esterni al rettangolo, si estendono fino ai valori estremi, minimo a sinistra (od in basso) massimo a destra (od in alto) e sono detti **whisker** (cioè baffi).

Nel **box plot modificato**, i “baffi” sono estesi fino ad un valore pari ad un multiplo opportuno della differenza interquartile (spesso pari a cinque volte), cosicché le osservazioni esterne, rispetto a questi limiti, sono qui assunte come outliers.

L'impiego dei grafici box plot e box plot modificato è utile per confronti immediati, tra i dati disponibili, poiché di facile comprensione.

#### 2.1.5 Distribuzioni di probabilità

Un campione è un insieme di elementi estratti, a caso, da una popolazione più ampia ed una distribuzione di probabilità è un modello matematico, collegare il valore della variabile alla probabilità che questo valore si trovi all'interno della popolazione (Figura 2.1.2).

- ❑ Quando il parametro da misurarsi può assumere solo determinati valori (ad esempio, come i numeri interi), la distribuzione di probabilità è definita una **distribuzione discreta** e si presenta come una serie di segmenti di altezza proporzionale alla probabilità. Alcune distribuzioni discrete compaiono frequentemente nel controllo statistico della qualità, come la distribuzione binomiale, la distribuzione poissoniana, la distribuzione ipergeometrica e la distribuzione binomiale negativa.
- ❑ Quando la variabile da misurarsi è espressa con una scala continua, la sua distribuzione di probabilità è definita una **distribuzione continua** e si presenta come una curva regolare, avente l'area, sottostante la curva, uguale alla probabilità. Anche alcune distribuzioni continue compaiono frequentemente nel controllo statistico della qualità, come la distribuzione normale (dove a sostegno della sua centralità è fondamentale il teorema del limite centrale), la distribuzione lognormale, la distribuzione esponenziale e la distribuzione gamma.

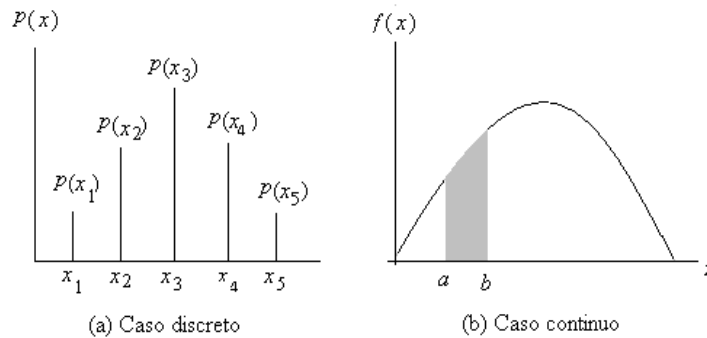


Fig. 2.1.2 – Distribuzioni di probabilità

## 2.4 Grafici di probabilità

La rappresentazione detta grafico di probabilità è un metodo euristico per determinare, se i dati campionari sono conformi a una distribuzione ipotizzata, con un semplice esame, puramente soggettivo, dei dati. La procedura generale è molto semplice e può essere eseguita velocemente, in quanto la rappresentazione usa una speciale rappresentazione grafica, detta carta di probabilità, realizzata per la specifica distribuzione ipotizzata (con lo scopo di mettere in corrispondenza lineare ascisse ed ordinate).

Per costruire il grafico di probabilità occorre dapprima ordinare le osservazioni disponibili, dalla più piccola alla più grande. Se la distribuzione di probabilità ipotizzata descrive adeguatamente i dati, i punti disegnati cadono approssimativamente lungo una linea retta; se invece i punti disegnati deviano, significativamente e sistematicamente, da una linea retta, il modello ipotizzato non è appropriato.

I grafici di probabilità sono estremamente utili e sono spesso la prima tecnica esplorativa, usata per la scelta della distribuzione di probabilità più adatta a descrivere i dati. Nell'usare i grafici di probabilità, la decisione è presa, di solito, attraverso una valutazione, puramente soggettiva, del grafico di probabilità. Insieme ai grafici di probabilità, si possono usare anche opportuni test statistici, circa la bontà dell'adattamento.

## 3. Inferenze riguardanti la qualità dei processi

In tutti gli esempi di statistica descrittiva, si assume che, impiegando le leggi di distribuzione, per modellare o descrivere le manifestazioni di un processo, i parametri delle distribuzioni ed i parametri dei processi siano noti. In generale, questa assunzione è non tuttavia realistica; infatti se si considera, ad esempio, l'uso della distribuzione binomiale, per rappresentare il numero di unità non conformi trovate, in un campione ottenuto da un processo, si assume che sia noto il parametro  $p$  della distribuzione.

Al contrario, questo parametro è da interpretarsi concretamente come la frazione di unità, non conformi, presenti nel processo, che risulta impossibile conoscere, con esattezza, relativamente ad un processo effettivo. Inoltre se il vero valore  $p$  è davvero noto e costante nel tempo, l'attività di monitoraggio e le procedure di controllo sono inutili.

Allora si può considerare che, in generale, i parametri del processo sono ignoti e possono variare, nel tempo, rendendo necessaria l'introduzione di una procedura, per stimare i parametri delle distribuzioni di probabilità coinvolte e risolvere ogni altro problema di inferenza o decisione, a riguardo. Si dimostrano così utili i metodi statistici e le tecniche di stima e verifica delle ipotesi. Pertanto queste tecniche costituiscono il substrato necessario per molte metodologie impiegate nel controllo statistico della qualità, proprio perché i metodi statistici di base dell'inferenza sono utili anche nell'azione di miglioramento della qualità. Gli strumenti

chiave comprendono le stime, puntuali ed intervallari, di medie e varianze delle distribuzioni di probabilità, dei parametri delle leggi binomiali, le verifiche di ipotesi, riguardanti le stesse grandezze, ed inoltre l'impiego delle carte di probabilità normali.

### 3.1 Statistiche e distribuzioni campionarie

Lo scopo dell'inferenza statistica è trarre conclusioni od assumere decisioni riguardanti una popolazione, sulla base di un campione estratto dalla popolazione stessa. In queste analisi, spesso si utilizzano campioni casuali, dove la parola casualità è riferita ad un metodo o selezione campionaria che eviti criteri sistematici.

- ❑ Nei campioni estratti da popolazioni costituite da infinite unità o da popolazioni finite, dove il campionamento è effettuato con rimessa o reintroduzione, le osservazioni sono selezionate in modo indipendente l'una dalle altre.
- ❑ Nel campionamento da una popolazione finita, senza reintroduzione, costituita da  $N$  unità, il campione di  $n$  unità è invece un campione casuale, se ciascuno dei possibili campioni ha un'eguale probabilità di estrazione, essendo il numero dei possibili campioni pari a:

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

Molti dei metodi utilizzati assumono che il campionamento sia casuale, benché esistano altre strategie di campionamento, impiegate utilmente nell'ambito del controllo di qualità. Attenzione particolare deve essere rivolta ai metodi di analisi, in modo che siano congrui con il piano di campionamento; infatti le procedure inferenziali, valide per campioni casuali, possono comportare errori rilevanti, quando siano applicate in presenza di tecniche di campionamento differenti.

Se si conosce la legge di distribuzione della popolazione della quale è estratto il campione, è possibile determinare la legge di distribuzione (detta **distribuzione campionaria**) delle diverse statistiche, ottenute sui dati campionari.

### 3.2 Stima puntuale dei parametri del processo

Una variabile casuale è caratterizzata dalla legge di distribuzione di probabilità e questa legge è identificata dai suoi **parametri** (ad esempio la media e la varianza sono i parametri della distribuzione normale, ad una dimensione). Nell'ambito del controllo statistico della qualità, la funzione di distribuzione è impiegata per descrivere o rappresentare, mediante un modello qualche caratteristica della qualità, come una grandezza od una dimensione critica del prodotto, oppure la frazione di unità difettose nel processo di fabbricazione. Ne consegue che interessa effettuare inferenza, relativamente ai parametri delle distribuzioni di probabilità di queste grandezze e, poiché questi parametri non sono noti, in generale, si richiedono procedure, per stimarli sulla base dei dati campionari.

### 3.3 Inferenza statistica per un singolo campione

Le tecniche di inferenza statistica possono essere distinte in due grandi categorie: la stima parametrica ed i test d'ipotesi.

Nella stima puntuale dei parametri di un processo, un'**ipotesi statistica** è un'affermazione circa i valori dei parametri di una distribuzione di probabilità, contrapposta ad un' **ipotesi alternativa** che, in relazione al problema specifico, possono essere differenti (bilaterali od unilaterali).

Le procedure di verifica dell'ipotesi risultano di rilevante utilità nei problemi di controllo statistico della qualità e costituiscono la base di differenti tecniche e procedure, impiegate nell'ambito dello studio statistico del controllo di processo.

### 3.3.1 L'impiego del *P*-value nelle verifiche d'ipotesi

Il modo consueto per presentare i risultati di una verifica d'ipotesi è quello di indicare che l'ipotesi nulla sia o non sia rifiutata, rispetto ad un preassegnato valore del livello di significatività. Queste conclusioni possono spesso risultare inadeguate, perché non danno alcuna idea del valore ottenuto della statistica del test, non esprimendo, se essa sia di poco entro la regione di rifiuto, oppure se si trovi notevolmente all'interno della regione stessa. Inoltre questa modalità impone un livello di significatività predefinito, anche per altri fruitori delle informazioni.

Di conseguenza, questo approccio può essere insoddisfacente, per il responsabile delle decisioni che può trovare complesso, confrontare il risultato, con tutti i rischi che implica il valore del livello di significatività. Per superare queste difficoltà è ampiamente adottato, nella pratica, l'approccio del ***P*-value**. Con il termine *P*-value (o valore *P*), si intende la probabilità che la statistica test assuma un valore, almeno come estremo, uguale a quello osservato, quando l'ipotesi fondamentale o nulla è vera. In questo modo, il *P*-value contiene molte informazioni, relative all'evidenza di non validità di questa ipotesi, così il responsabile delle decisioni può trarre una conclusione, per ogni specifico livello di significatività.

**Definizione formale di *P*-value:** il *P*-value è il più piccolo livello di significatività che conduce al rifiuto dell'ipotesi fondamentale o nulla.

E' abitudine chiamare una statistica test significativa, quando l'ipotesi nulla è rifiutata. Allora il *P*-value è il più piccolo livello di significatività per cui i dati osservati sono significativi. Pertanto disponendo del *P*-value, il responsabile delle decisioni può stabilire, da solo, quanto siano significativi i dati osservati, senza che chi è predisposto all'analisi dei dati debba imporre un livello di significatività, a priori.

Non è sempre facile ottenere un esatto *P*-value, per ogni procedura di verifica, comunque molti programmi, dedicati all'analisi statistica, presentano il *P*-value. Inoltre è possibile impiegare tabelle statistiche che approssimano i *P*-value, per alcuni casi specifici.

## 4. Il controllo statistico di processo

E' difficile ispezionare o verificare la qualità di un prodotto; piuttosto è meglio avere un prodotto che sia qualitativamente accettabile, già in fase di produzione. Questo richiede che il processo sia stabile e che le persone impiegate nella produzione siano continuamente impegnate, anche nel miglioramento del processo e nella riduzione della variabilità dei fattori coinvolti nella produzione. L'**SPC (Statistical Process Control)** è lo strumento primario, per conseguire questo risultato, e le **carte di controllo** sono lo strumento più semplice, per definire una procedura di controllo statistico di processo. A riguardo, gli scopi fondamentali delle carte di controllo sono tre:

- riduzione della variabilità del processo;
- monitoraggio e controllo del processo;
- stima dei parametri del prodotto e del processo,

laddove la teoria ed i metodi del controllo statistico di un processo riguardano:

- gli strumenti statistici di base del SPC, noti anche con il nome di “magnifici sette”;
- le basi statistiche delle carte di controllo di Shewhart;
- alcuni problemi pratici, incontrati nell’applicazione dell’SPC.

#### 4.1 I “magnifici sette”

L’SPC è applicato a qualsiasi processo ed i sette più importanti strumenti statistici di cui si avvale sono:

- istogrammi e grafici “rami e foglie”;
- fogli di controllo;
- grafici di Pareto;
- diagrammi causa ed effetto;
- diagrammi sulla concentrazione dei difetti;
- grafici a dispersione;
- carte di controllo.

#### 4.2 Fonti di variabilità nella qualità

Ogni processo è sempre soggetto ad una certa variabilità intrinseca o naturale. Questa variabilità naturale o rumore di fondo è il risultato dell’effetto cumulato di molti piccoli, ma ineliminabili, fattori costanti o casuali. Un processo la cui variabilità sia provocata solo da fattori casuali è detto sotto controllo, mentre le fonti di variabilità, non ricondotte a fattori casuali, sono dette “fattori specifici” ed un processo, funzionante in presenza di fattori specifici è detto fuori controllo. In generale, i processi operano in situazione di controllo, producendo pezzi di qualità accettabile, per lunghi periodi di tempo.

Tuttavia possono verificarsi fattori specifici, apparentemente casuali, così da comportare la produzione di grandi quantità di pezzi non conformi agli standard qualitativi. Infatti quando il processo è sotto controllo, la maggior parte dei valori della grandezza oggetto di controllo cade tra i **limiti di specifica** superiore ed inferiore (indicati con USL: Upper Specific Limit, e LSL: Lower Specific Limit). Invece, quando il processo è fuori controllo, un elevato numero di determinazioni campionarie cade fuori dai limiti di queste specifiche di controllo.

L’obiettivo primario del controllo statistico di un processo è individuare, il più velocemente possibile, il verificarsi di fattori specifici. Infatti quanto più veloce è l’individuazione delle cause, tanto prima si possono avviare azioni di correzione, così da evitare la produzione di molti pezzi di qualità non accettabile. Le carte di controllo sono uno strumento ampiamente usato, per questi scopi e per determinare la capacità del processo stesso. Infine si ricordi che lo scopo del controllo statistico di un processo è eliminare la variabilità all’interno del processo stesso e, per quanto non sia possibile eliminarla completamente, le carte di controllo costituiscono un efficace strumento, per ridurla il più possibile.

### 4.3 Fondamenti statistici delle carte di controllo

Una tipica carta di controllo è riportata nella Figura 1.2.1 e descrive una certa qualità di un prodotto misurata in diversi istanti temporali. La carta riporta una linea centrale che rappresenta il valore medio della qualità, in generale, corrispondente al valore desiderato, quando il processo è sotto controllo. Le altre due linee orizzontali sono dette limite di controllo superiore (UCL) e limite di controllo inferiore (LCL).

Questi limiti di controllo sono scelti, in modo tale che, se il processo è sotto controllo, quasi tutti i valori campionari cadono al loro interno e nessun intervento correttivo sul processo è necessario. Invece se più punti cadono al di fuori di questi limiti, questi devono essere interpretati come una spia del fatto che il processo è fuori controllo e si rendono necessarie azioni correttive o di indagine, sul processo stesso, per individuare ed eliminare le cause che hanno portato all'insorgere di quei fattori specifici. Si è poi soliti unire tutti i punti campionari consecutivi con tratti continui, così da rendere più facile la visualizzazione dell'evoluzione del processo nel tempo.

Esiste uno stretto legame fra carte di controllo e test d'ipotesi; a riguardo, posto l'asse verticale di una carta come media campionaria. Se il valor osservato cade tra i limiti di controllo, la media del processo è sotto controllo, cioè pari a un valore ad dato valore di riferimento. D'altra parte, se il valor osservato supera uno dei limiti di controllo, la media del processo è fuori controllo, ovvero è pari ad un valore di riferimento, diverso (ed inaccettabile).

Supponendo uno schema generico di carta di controllo, sia data una qualche statistica campionaria  $\omega$  che misura una certa caratteristica di un prodotto e si supponga che la sua media sia  $\mu_\omega$  e la sua deviazione standard sia  $\sigma_\omega$ , cosicché UCL, la linea centrale (CL) e LCL siano:

$$UCL = \mu_\omega + L\sigma_\omega$$

$$CL = \mu_\omega$$

$$LCL = \mu_\omega - L\sigma_\omega$$

dove  $L$  è la distanza dei limiti di controllo della linea centrale, espressa in unità di deviazione standard. Le carte di controllo, costruite secondo questi criteri, sono chiamate **carte di controllo di Shewhard**.

La carta di controllo è uno strumento per descrivere in maniera sistematica quello che è chiamato controllo statistico e, di conseguenza, può essere usata per il controllo di sorveglianza on-line: i dati campionari sono raccolti periodicamente e rappresentati, nella carta di controllo; se il valore osservato cade tra i limiti di controllo e se, nel tempo, non evidenzia andamenti non casuali diremo che il processo è sotto controllo. L'uso più importante è tuttavia quello di contribuire al miglioramento del processo.

Le carte di controllo possono essere usate in due modi, a seconda della caratteristica della variabile oggetto di studio. Se la caratteristica di un prodotto è rappresentabile su una scala continua di valori, è chiamata variabile ed è possibile descriverla, come una misura di centralità ed una di variabilità. Le carte di controllo, per la centralità e la variabilità di un processo, sono dette **carte di controllo per variabili**.

D'altra parte, molte caratteristiche dei prodotti non possono essere misurate su scala continua, né su scale genericamente quantitative. In questi casi, ciascuna unità prodotta è valutata conforme a seconda che possieda (o meno) certi attributi, od a seconda del numero dei difetti, presenti nell'unità prodotta. Le carte di controllo, costruite sulla base di queste grandezze, sono dette **carte di controllo per attributi**.

### **4.3.1 Scelta dei limiti di controllo**

La definizione dei limiti di controllo è uno dei passaggi più critici, nella progettazione di una carta di controllo. Infatti più sono collocati lontano dalla linea centrale, minore è il rischio che un punto si posizioni al di fuori dei limiti, indicando una situazione di fuori controllo, quando invece nessun fattore si manifesta (rischio primo tipo). D'altra parte, ampliare i limiti di controllo comporta il rischio che un punto cada all'interno dei limiti di controllo, quando il processo non è, di fatto, sotto controllo (rischio secondo tipo). Invece se i limiti sono avvicinati alla linea di controllo, si ha l'effetto opposto, così il rischio di primo tipo aumenta, mentre quello di secondo tipo diminuisce.

Ad esempio, per la carta in Figura 1.2.1, se si ipotizza che il diametro della fascia elastica del pistone abbia distribuzione normale e che i limiti controllo siano a 3-sigma, utilizzando la tavola della variabile casuale normale standardizzata, si trova che l'errore di primo tipo è 0.0027, ovvero un errato segnale di fuori controllo (o falso allarme) è osservato, in probabilità, 27 volte ogni 10.000 campioni. Inoltre la probabilità che un punto superi uno dei limiti, posti a 3-sigma, quando il processo è sotto controllo è solo 0.00135. Questi limiti di controllo sono detti limiti con probabilità 0.001, dove in generale, la scelta dei limiti a 3-sigma dà buoni risultati nelle applicazioni.

### **4.3.2 Limiti di sorveglianza delle carte di controllo**

Alcuni analisti suggeriscono di utilizzare sulla stessa carta di controllo due differenti limiti come quelli evidenziati nell'immagine. I limiti esterni, ad esempio, posizionati a 3-sigma, sono gli usuali limiti operativi e, quando un punto cade al di fuori di questi, deve essere intrapresa una ricerca delle cause e devono essere predisposti gli eventuali interventi correttivi. I limiti interni, spesso posizionati a 2-sigma, sono chiamati limiti di sorveglianza superiore (UWL: Upper Warning Limits) e inferiore (LWL: Lower Warning Limits).

Se uno o più punti cadono tra i limiti di controllo e limiti di sorveglianza od anche solo in prossimità dei limiti di sorveglianza, si deve ritenere che il processo non stia funzionando correttamente. Per essere più certi di questa congettura e senza dover intraprendere azioni correttive sul processo, si è soliti aumentare la frequenza di campionamento e/o la dimensione del campione, cosicché molte più informazioni siano analizzate, nell'intorno temporale dell'istante in cui il problema sembra essersi manifestato. Questi interventi sono detti schemi adattivi o schemi di campionamento con dimensione campionaria variabile.

### **4.3.3 Dimensione del campione e frequenza di campionamento**

Nel progetto di una carta di controllo, bisogna specificare la dimensione del campione da analizzare e la frequenza di campionamento. In generale, quanto più grande è il campione, tanto più facile è individuare piccoli spostamenti o regolazioni del processo. Infatti rappresentata sulla carta di controllo la curva operativa caratteristica, per diverse le dimensioni campionarie della variabile oggetto, per scegliere la dimensione campionaria ottimale, bisogna conoscere quale sia lo scostamento del processo che si vuole individuare più velocemente.

Contemporaneamente si deve determinare la frequenza di campionamento. La situazione ottimale è poter esaminare grandi campioni di frequente, ma questa è una situazione sicuramente poco accettabile, dal punto di vista economico. Si tratta così di un problema di buona allocazione degli sforzi di campionamento, dovendosi esaminare piccoli campioni, di frequente, o grandi campioni, ad intervalli più distanziati.

Due strumenti utili per calcolare la dimensione campionaria e la frequenza di campionamento ottimali sono la **lunghezza media delle frequenze** (ARL: *Average Run Length*) e il **tempo medio al segnale** (ATS: *Average Time to Signal*). Data la probabilità  $p$  che un punto superi i limiti di controllo, l'ARL è il numero medio di punti da osservare, prima che un punto cada fuori dai limiti di controllo.

$$ARL = 1/p$$

Invece il tempo medio al segnale ATS è il prodotto dell'ARL per l'intervallo medio di tempo, intercorrente tra due campioni, ed indica il tempo medio intercorrente tra due segnali di fuori controllo. Se i campioni sono esaminati, ad intervalli di tempo costante  $h$ , si ha:

$$ATS = h \cdot ARL$$

#### 4.3.4 Sottogruppi razionali

I sottogruppi razionali sono quei campioni, scelti in modo che, se sono presenti fattori specifici, la probabilità di osservare differenze tra i campioni sia massimizzata, mentre la probabilità di osservare differenze tra le unità che compongono il campione, a causa di fattori specifici, sia minimizzata. Due approcci permettono la costruzione di sottogruppi razionali.

- Il primo suggerisce di costruire i campioni, in modo che contengano unità prodotte nello stesso istante temporale. Questo approccio deve essere seguito, quando lo scopo è di usare la carta di controllo per individuare scostamenti sistematici della produzione, perché minimizza la probabilità di avere variabilità attribuibili a fattori specifici, all'interno del campione, mentre massimizza la probabilità di avere elevata variabilità tra i campioni, se presentano scostamenti sistematici dei parametri, oggetto di controllo. Inoltre questo approccio fornisce una migliore stima della deviazione standard del processo, se la carta di controllo è per variabili.
- Il secondo approccio consiste nella costruzione di campioni, rappresentativi della produzione intercorsa dall'ultimo controllo effettuato, cosicché ciascun sottogruppo è un campione casuale della produzione ottenuta nell'intervallo di campionamento. Questo approccio è usato, quando si ipotizza che il processo possa uscire di controllo, nell'intervallo di tempo intercorso dall'ultimo campione esaminato.

#### 4.3.5 Analisi degli andamenti tipici di una carta di controllo

Una carta di controllo può indicare una situazione di fuori controllo, quando uno o più punti cadono oltre i limiti di controllo, oppure quando l'andamento descritto dai punti, pur interni ai limiti, non è casuale. Infatti la non casualità dei punti è sintomo di un processo che produce pezzi le cui misure di qualità possono essere descritte da un modello e, come tali, prevedibili. Invece per essere definito sotto controllo, un processo deve, in generale, avere la proprietà di produrre pezzi, con caratteristiche aventi distribuzione casuale, secondo un modello probabilistico definito.

Una successione di punti avente un andamento crescente o decrescente è detta sequenza, così come è detta sequenza di lunghezza 2: una successione di punti, alternativamente uno sopra ed uno sotto la linea centrale. In generale, sequenze di lunghezza maggiore di 7 hanno una probabilità di verificarsi molto bassa, essendo di tipo non casuale, e così sono spesso un segnale di fuori controllo.

L'abilità nell'individuazione di un comportamento anomalo e la sua attribuzione a fattori specifici richiede esperienza e profonde conoscenze dei processi. In particolare, alcune regole decisionali, per l'individuazione di comportamenti anomali, stabiliscono che un processo è da considerare fuori controllo, se:

- un punto cade al di fuori dei limiti di 3-sigma;
- due punti, su tre consecutivi, cadono oltre i limiti di sorveglianza, posizionati a 2-sigma;
- quattro punti, su cinque consecutivi, cadono oltre la distanza di 1-sigma dalla linea centrale;
- otto punti consecutivi cadono tutti dalla stessa parte della linea centrale.

Queste regole si applicano ad una sola parte della carta di controllo e così, se un punto cade sopra l'UWL ed il successivo sotto l'LWL, questo non indica una situazione di fuori controllo.

#### 4.3.6 Commento alle regole di sensibilità per le carte di controllo

Esistono diversi criteri, per stabilire, se un processo è fuori controllo. Il criterio di base è che uno o più punti cadano al di fuori dei limiti di controllo. Tuttavia in generale, sono usati criteri aggiuntivi, per aumentare la sensibilità di una carta, nel rilevare regolazioni del processo, anche piccole e, se solo uno di questi criteri è soddisfatto, si considera il processo fuori controllo.

Segnale di intervento standard	Uno o più punti cadono al di fuori dei limiti di controllo
	Due punti, su tre consecutivi, cadono oltre i limiti di sorveglianza, posizionati a 2-sigma (ma rimangono entro i limiti a 3-sigma)
	Quattro punti, su cinque consecutivi, cadono oltre la distanza di 1-sigma dalla linea centrale
	Otto punti consecutivi cadono dalla stessa parte della linea centrale
	Sei punti consecutivi sono in ordine crescente o decrescente
	Quindici punti consecutivi sono in una definita zona (sotto o sopra la linea centrale)
	Quattordici punti consecutivi si alternano a zig-zag
	Otto punti consecutivi si alternano attorno alla linea centrale (ma nessuno è nella zona definita)
	Si manifesta un comportamento non casuale dei dati
	Uno o più punti si posizionano vicino ai limiti di sorveglianza e di controllo

In ogni caso, le regole, riassunte nella tabella, devono essere usate, con una certa cautela, se non si vuole andare incontro ad un numero eccessivo di falsi allarmi.

#### 4.3.7 Prima e seconda fase dell'applicazione delle carte di controllo

Nella **prima fase**, si esegue un'analisi preliminare del processo, raccogliendo ed analizzando un insieme retrospettivo di dati del processo, al fine di determinare, se rappresentano una situazione in stato di controllo e costruire i limiti di controllo, attendibili per monitorare correttamente i dati futuri. Allora nella prima fase, le carte di controllo si utilizzano, per isolare sequenze di dati del processo, in stato di controllo statistico.

Nella prima fase sono stabiliti dei limiti preliminari, basandosi sui sottogruppi disponibili e su tutti i dati, rappresentati sulle carte di controllo. I punti fuori dai limiti di controllo sono esclusi e si calcolano i nuovi limiti di controllo, successivamente sono esaminati altri dati e comparati con questi limiti aggiornati. Talvolta questa fase dell'analisi deve essere ripetuta parecchie volte, fino ad ottenere una serie di dati del processo

stabilizzato che si può ritenere capace di rappresentare bene l'andamento di un processo in controllo. In generale, la carte di controllo di Sheward sono molto efficienti nella prima fase e sono facili da costruire ed interpretare. Infatti sono efficienti nell'individuare la presenza di forti scostamenti di livello dei parametri del processo, come pure di outliers (singole escursioni, derivate da cause assegnabili di breve durata, errori di misurazione e/o registrazione dei dati, e così via).

Nella **seconda fase**, si usano poi le carte di controllo, per monitorare il processo, confrontando la statistica campionaria di ogni campione, estratto dal processo, con i limiti di controllo, stabiliti nella prima fase. Nella seconda fase, si assume usualmente che il processo sia in condizioni di stabilità. Spesso le cause assegnabili che si verificano, nella seconda fase, sono costituite da limitati spostamenti di livello, poiché la maggior parte delle fonti di variabilità, veramente dannose, devono essere individuate durante la prima fase e le loro cause rimosse, anche per il prosieguo. Pertanto l'enfasi è, in questa fase, posta sul monitoraggio del processo e non sul suo riassetto. verso lo stato di controllo.

#### **4.4 Altri strumenti dei “magnifici sette”**

Sebbene la carta di controllo sia un strumento potente, per il controllo ed il miglioramento della qualità, di fatto, risulta più efficace, se è usata in un contesto, dove si ha un'ampia integrazione degli strumenti di SPC. Infatti questi “magnifici sette” devono essere usati, diffusamente e quotidianamente, ad ogni livello.

##### Fogli di controllo

I primi passi per l'implementazione dell'SPC, richiedono di provvedere ad una raccolta di dati, riguardanti il processo oggetto di controllo e, a questo scopo, uno strumento utile è il foglio di controllo. A riguardo, nel progetto di un foglio di controllo, è importante definire chiaramente il tipo di informazione raccolta, la data, l'operatore che effettua la raccolta e qualsiasi altra nota, per rendere più intelligibile la provenienza del dato.

##### Grafico di Pareto

Uno degli strumenti più utili nell'SPC riporta l'istogramma ed associata la distribuzione di frequenza cumulata di dati qualitativi, distinti per categoria ed ordinati per frequenza. Infatti con il grafico di Pareto, l'utente può facilmente individuare la più frequente tipologia di difetti e, in questo modo, le cause più spesso manifeste, anche se non quelle più importanti, per il funzionamento del processo.

##### Diagramma causa ed effetto

Una volta identificato ed isolato un errore, un problema od un elemento difettoso, devono essere cercate le cause potenziali di questo effetto indesiderato, costruendo un diagramma causa ed effetto per passi:

- definire il problema o l'effetto da analizzare;
- preparare il team per eseguire l'analisi;
- predisporre la casella dell'effetto ed una linea centrale;
- specificare le tipologie delle più importanti cause ed unirle, a tratti, come caselle connesse alla linea centrale;
- indicare le possibili cause e classificarle nelle tipologie individuate al passo precedente e, se necessario, creare altre tipologie;

- ordinare le cause, in base alla maggiore o minore probabilità che possano creare il problema oggetto di analisi.

Quanto più dettagliati sono i diagrammi di causa ed effetto, tanto più efficaci si dimostrano nell'aiutare nella soluzione del problema.

#### Diagrammi sulla concentrazione dei difetti

In un particolare disegno sulla struttura del prodotto, sono riportate tutte le visuali rilevanti del prodotto stesso. I diversi tipi di difetti sono evidenziati su questo disegno, con lo scopo di verificare, se la loro collocazione può costituire un'utile informazione sulle cause potenziali del difetto rilevato. Una volta raccolti i diagrammi sulla concentrazione dei difetti, per un numero sufficientemente elevato di unità prodotte, spesso si riesce a individuare alcune tipologie che aiutano a capire le cause e gli effetti.

#### Grafici a dispersione

Per individuare potenziali relazioni funzionali tra due variabili, si considerano due variabili e, di queste, sono raccolte più coppie di dati, rappresentate poi su un piano cartesiano, cosicché spesso l'andamento del grafico a dispersione indica quale relazione sussiste tra le due variabili.

### **5. Carte di controllo per variabili**

Le carte di controllo per caratteristiche qualitative, misurate su scala numerica, come ad esempio la lunghezza, l'ampiezza, la temperatura, il volume, ecc. sono dette carte di controllo per variabili.

#### **5.1 Introduzione**

Una caratteristica misurabile, come una dimensione, il peso od il volume, è chiamata variabile e, in un contesto di SPC, è assolutamente necessario, per poter controllare la media e la variabilità, monitorare con carte  $\bar{x}, R$  o  $\bar{x}, S$  (dove  $R$  e  $S$  sono rispettivamente il range e la deviazione standard campionari).

#### **5.2 Carte di controllo $\bar{x}, R$**

Dati una caratteristica quantitativa  $X$ , distribuita secondo la legge di una variabile casuale normale, con media  $\mu$  e deviazione standard  $\sigma$ , ipotizzate note, ed un suo campione, di dimensione  $n$ , si può calcolare il valore medio del campione  $\bar{x}$  che si dimostra avere una distribuzione normale, con la stessa media ed una deviazione standard:  $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ . Allora la probabilità che un'altra media, proveniente da un campione di pari dimensione, cada fra:

$$\mu \pm Z_{\alpha/2} \sigma_{\bar{x}} = \mu \pm Z_{\alpha/2} \sigma / \sqrt{n}$$

è pari ad  $1 - \alpha$ . Pertanto considerati noti i valori teorici, della variabile casuale, e sostituendo  $Z_{\alpha/2}$  con il numero 3, si possono utilizzare i limiti di controllo a 3-sigma e le suddette equazioni interpretare come limiti di controllo, superiore ed inferiore, di una carta di controllo per la media campionaria.

In questo modo, se una media campionaria cade al di fuori di questi limiti, è un segnale che la media del processo non è più pari a  $\mu$ .

Di fatto, l'ipotesi introdotta in precedenza di ritenere noti i valori teorici, della variabile casuale, non è quasi mai soddisfatta. Di conseguenza, questi parametri devono essere stimati sulla base di un certo numero di campioni preliminari (in generale, almeno  $20 \div 25$ ), opportunamente estratti, in un periodo in cui il processo è ritenuto sotto controllo. A tal fine, si suppone disponibili  $m$  campioni, ciascuno con  $n$  determinazioni casuali della caratteristica qualitativa, oggetto di interesse (tipicamente  $n$  è un numero piccolo, spesso pari a 4, 5 o 6 ed una, così piccola, dimensione campionaria è giustificata, se si utilizzano sottogruppi razionali o se sono piuttosto alti i costi di campionamento ed ispezione, associati alle variabili quantitative). Dopodiché calcolate le medie di ciascun campione, se la media  $\mu$  del processo non è nota, il miglior stimatore è la media delle medie  $\bar{\bar{x}}$ , degli  $m$  campioni, da usare come linea centrale, nella carta di controllo.

Tuttavia per costruire i limiti di controllo, occorre anche una stima della deviazione standard  $\sigma$  del processo. A tal fine, date le dimensioni, così piccole, dei campioni, è possibile stimare  $\sigma$ , usando i range degli  $m$  campioni, dove ciascun range è la differenza tra la più grande e la più piccola determinazione campionaria:

$$R = x_{max} - x_{min}$$

cosicché lo stimatore del range del processo, da usare per costruire i limiti di controllo, è il range medio  $\bar{R}$  e, tramite questo, si hanno infine le seguenti carte di controllo:

- limiti di controllo per la carta di controllo  $\bar{x}$  (dove la costante  $A_2$  è tabulata per valori di  $n$ ):

$$UCL = \bar{\bar{x}} + A_2 R$$

$$CL = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - A_2 R$$

- limiti di controllo per la carta di controllo  $R$  (dove le costanti  $D_4$  e  $D_3$  sono tabulate per valori di  $n$ ):

$$UCL = D_4 R$$

$$CL = R$$

$$LCL = D_3 R$$

In questo modo, anche la variabilità del processo può essere monitorata, riportando i valori del range, di ogni campione, su una carta di controllo  $R$ .

### 5.3 Carte di controllo $\bar{x}, S$

In presenza di dimensioni campionarie sufficientemente grandi, è opportuno stimare la deviazione standard del processo con la deviazione standard campionaria  $S$  e non con il range campionario  $R$ , ottenendo così le carte di controllo  $\bar{x}, S$ , preferibili alle corrispondenti carte di controllo  $\bar{x}, R$ , quando la dimensione campionaria è abbastanza grande ( $n \geq 10 \div 12$ ), oppure la dimensione campionaria è variabile.

La costruzione delle carte di controllo  $\bar{x}, S$  richiede gli stessi passi necessari per la costruzione delle carte di controllo  $\bar{x}, R$ , ad eccezione del fatto che, per ogni campione, si devono calcolare  $\bar{x}$  e  $S$ , utilizzati poi per costruire le corrispondenti carte.

Data la varianza incognita  $\sigma^2$  di una popolazione, un suo stimatore corretto è la varianza campionaria  $S^2$  (corretta per i gradi di libertà, pari a  $n - 1$ ), mentre la deviazione standard campionaria  $S$  non è uno stimatore corretto. D'altra parte, se la distribuzione oggetto di studio è normale, allora  $S$  è la stima di una grandezza pari a  $c_4\sigma$ , dove  $c_4$  è una costante tabulata per valori di  $n$ . Inoltre la deviazione standard di  $S$  è  $\sqrt{1 - c_4^2}$ . Pertanto dato il caso in cui a  $\sigma$  è assegnato il valore standard  $c_4\sigma$ , la linea centrale della carta di controllo ed i limiti di controllo a 3-sigma, per  $S$ , sono:

$$UCL = c_4\sigma + 3\sigma\sqrt{1 - c_4^2} = B_6\sigma \quad \text{con} \quad B_6 = c_4 + \sqrt{1 - c_4^2}$$

$$CL = c_4\sigma$$

$$LCL = c_4\sigma - 3\sigma\sqrt{1 - c_4^2} = B_5\sigma \quad \text{con} \quad B_4 = c_4 - \sqrt{1 - c_4^2}$$

Dopodiché quando sono noti i valori teorici della media  $\mu$  e della deviazione standard  $\sigma$  della popolazione, questi si possono usare per la costruzione della carta di controllo di  $\bar{x}$  i cui parametri sono:

$$UCL = \mu + 3\sigma/\sqrt{n} = \mu + A\sigma \quad \text{con} \quad A = 3/\sqrt{n}$$

$$CL = \mu$$

$$LCL = \mu - 3\sigma/\sqrt{n} = \mu - A\sigma$$

dove anche la costante  $A$ , per comodità, è tabulata per valori di  $n$ .

## 6. Carte di controllo per attributi

Molte caratteristiche relative alla qualità di un prodotto non possono essere rappresentate numericamente. In questi casi, si usa classificare ciascun oggetto esaminato solo come conforme (o non conforme), secondo certe caratteristiche qualitative prescelte, dette attributi, dove i termini difettoso (o non difettoso) sono spesso usati, per identificare questa classificazione.

### 6.1 Introduzione

Le carte di controllo per attributi non sono così informative come le carte di controllo per variabili. Infatti l'informazione contenuta in una misura è più informativa di quella ottenuta classificando una unità solo come conforme (o non conforme). In ogni caso le carte di controllo per attributi hanno importanti applicazioni, per migliorare la qualità, nei settori dove le grandezze caratteristiche non sono misurabili su scala numerica.

### 6.2 Carte di controllo per frazioni di non conformi

La frazione di non conformi è definita come il rapporto tra numero di unità non conformi, presenti in una popolazione, e numero di unità di quella popolazione.

Un'unità prodotta può avere diverse caratteristiche qualitative da esaminare simultaneamente e, se l'unità non soddisfa una o più caratteristiche è classificata come non conforme e la frazione campionaria di non conformi è definita dal rapporto fra il numero di unità non conformi, presenti nel campione, e la dimensione  $n$  del campione:

$$p = D/n$$

dove  $p$  è la probabilità di ottenere un'unità non conforme,  $n$  la dimensione del campione e  $D$  il numero di unità non conformi. La media e la varianza di  $p$  sono rispettivamente:

$$\mu = p \quad \text{e} \quad \sigma_p^2 = \frac{p(1-p)}{n}$$

### 6.2.1 Costruzione della carta di controllo

Supposta nota la frazione teorica  $p$  di entità non conformi, o comunque definito un valore standard, dalle equazioni delle carte di Shewhart (e dai loro fondamenti statistici), si definiscono la linea centrale della carta di controllo ed i limiti di controllo, per la frazione di non conformi (con valori di riferimento noti):

$$UCL = \mu_\omega + L\sigma_\omega = p + 3\sqrt{\frac{p(1-p)}{n}}$$

$$CL = \mu_\omega = p$$

$$LCL = \mu_\omega - L\sigma_\omega = p - 3\sqrt{\frac{p(1-p)}{n}}$$

dove  $L$ , di solito, pari a 3, è la distanza, espressa in unità di deviazione standard di  $\omega$ , dei limiti di controllo dalla linea centrale.

Per l'utilizzo della carta, bisogna estrarre campioni successivi di  $n$  unità, calcolare la frazione di non conformi  $p$  e rappresentare questa statistica sulla carta. Finché  $p$  rimane all'interno dei limiti di controllo e non si osserva alcun andamento anomalo (o non casuale) nella successione dei punti, si può affermare che il processo è sotto controllo a livello  $p$ . Invece se il punto cade al di fuori dei limiti (oppure se si osserva un andamento non casuale dei punti), si deve concludere che la frazione di non conformi ha probabilmente subito uno scostamento, ad un nuovo livello, ed il processo è fuori controllo.

Se la frazione teorica  $p$  di entità non conformi non è nota, deve essere stimata dai dati osservati. La procedura consiste nel selezionare  $m$  campioni preliminari, ciascuno di dimensione  $n$  (in generale, almeno  $20 \div 25$ ), opportunamente estratti, in un periodo in cui il processo è ritenuto sotto controllo.. Se si hanno  $D_i$  entità non conformi nel campione  $i$ -esimo, la corrispondente frazione campionaria di non conformi e la media di tutte le frazioni di non conformi sono rispettivamente:

$$p_i = D_i/n \quad i = 1, 2, \dots, m \quad \bar{p} = \frac{\sum_{i=1}^m D_i/n}{m} = \frac{\sum_{i=1}^m p_i}{m}$$

dove la statistica  $\bar{p}$  è lo stimatore della frazione incognita  $p$ .

In questo modo, si definiscono la linea centrale della carta di controllo ed i limiti di controllo, per la frazione di non conformi (con valori di riferimento non assegnati):

$$UCL = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$CL = \bar{p}$$

$$LCL = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

### 6.2.2. Dimensione campionaria variabile

Quando la dimensione campionaria è variabile, ad esempio, perché non si hanno sempre lo stesso numero di entità, anche i limiti di controllo di una carta per frazione di non conformi devono poter variare. Per gestire questa situazione, si possono perseguire tre criteri:

- limiti di controllo ad ampiezza variabile** (l'approccio più semplice che consiste nel calcolare i limiti per ogni ampiezza campionaria);
- limiti di controllo basati sulla dimensione campionaria media** (questo approccio consiste nel costruire i limiti di controllo, sulla base della dimensione campionaria media, ottenendo limiti di controllo approssimati);
- carta di controllo con valori standardizzati** (dove in presenza di  $n$  variabili, si possono usare valori standardizzati).

### 6.3 Carte di controllo per non conformità (numero dei difetti)

Un'unità non conforme è un prodotto non soddisfacente una o più caratteristiche qualitative. Ciascuna caratteristica qualitativa non soddisfatta è un difetto od una non conformità e, in generale, se un bene ha uno o più difetti, è non conforme. Tuttavia questo attributo dipende anche dalla gravità dei difetti, se questi non influiscono sul funzionamento dell'oggetto possono anche non essere tali da classificarlo necessariamente come non conforme, così come, da un altro punto di vista, pur non influenzando sul suo funzionamento, i difetti possono essere tali da rendere l'oggetto inutile, in dipendenza dal livello qualitativo prefissato. E' possibile costruire carte di controllo, per il numero totale di non conformità, per unità prodotta, e per il numero medio di non conformità, per unità prodotta.

#### 6.3.1 Procedure con dimensioni campionarie costanti

In generale, in una carta di controllo per non conformità, con i limiti di controllo a 3-sigma, il campione è costituito da un'unica unità di riferimento (che può essere anche una sola parte dell'intero prodotto); tuttavia si può anche scegliere più di un'unità prodotta, per costituire l'unità di riferimento.

- Carte di controllo per non conformità (con valori di riferimento assegnati):

$$UCL = c + 3\sqrt{c}$$

$$CL = c$$

$$LCL = c - 3\sqrt{c}$$

dove si ipotizza che  $c$  sia un valore noto e che, se  $LCL$  è negativo, questo è posto uguale a zero.

Se nessun valore di riferimento è assegnato, è possibile stimare  $c$  con il numero medio  $\bar{c}$  di difetti rilevati, in un campione preliminare, ed i limiti di controllo ottenuti si considerano solo limiti di prova delle carte di controllo per non conformità.

- ❑ Carte di controllo per non conformità (con valori di riferimento non assegnati):

$$UCL = \bar{c} + 3\sqrt{\bar{c}}$$

$$CL = \bar{c}$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}}$$

Le carte di controllo per non conformità sono talora indicate come carte  $c$ .

#### 6.4 Linee guida per l'applicazione delle carte di controllo

Quasi tutti i processi possono beneficiare dei metodi di SPC ed alcune linee guida, per l'applicazione delle carte di controllo, sono così riassunte.

- ❑ determinare quale caratteristica controllare;
- ❑ determinare dove le carte dovrebbero essere applicate nel processo;
- ❑ scegliere l'appropriata carta di controllo;
- ❑ attivare interventi migliorativi del processo, a fronte di risultati di analisi di SPC;
- ❑ scegliere quali strumenti impiegare, per raccogliere i dati per l'analisi di SPC.

##### Determinare quale caratteristica controllare e dove applicare le carte

Queste scelte non sono facili ed alcune linee guida sono solo proposte.

- ❑ All'inizio di un nuovo processo o di una nuova applicazione di SPC, le carte dovrebbero essere applicate a tutte le caratteristiche del prodotto, ritenute maggiormente importanti e, solo nel seguito, si stabilisce dove effettivamente sono necessarie.
- ❑ Le carte, ritenute non necessarie, devono essere eliminate ed eventualmente se ne aggiungono altre, se gli operatori lo ritengono necessario.
- ❑ Le informazioni sulle carte di controllo devono inizialmente essere tenute separate e, nel tempo, il loro numero deve diminuire, in relazione alla progressiva stabilizzazione del processo.
- ❑ In generale, con il passare del tempo, se si osserva che le carte sono utili per la produzione, si nota un aumento del numero delle carte per variabili ed una riduzione delle carte per attributi.
- ❑ Spesso si osserva l'applicazione delle carte, quando ormai il prodotto è finito; tuttavia è meglio applicarle il più presto possibile, per non incorrere in difetti fatali, in seguito, per l'intero prodotto.

- Le carte devono essere disponibili in luoghi il più vicino possibile alle attività, così da poter garantire un rapido intervento, in caso di una segnalazione di guasto; inoltre la vicinanza alla produzione consente di verificare velocemente gli effetti prodotti, da eventuali aggiustamenti sul processo.

#### Scegliere la carta appropriata

- Le carte  $\bar{x}, R$  o  $\bar{x}, S$  devono essere usate per misurare variabili, nei seguenti casi:
  - un nuovo processo produttivo sta per essere avviato su un processo già esistente;
  - il processo ha operato per un certo arco di tempo, ma ora presenta vari problemi di funzionamento;
  - il processo presenta problemi di funzionamento e la carta serve come strumento di diagnostica;
  - il processo è fuori controllo, oppure è sotto controllo, ma la difettosità è inaccettabile;
  - i controlli sono distruttivi o molto costosi;
  - si cerca di ridurre il numero di accettazione, quando il processo è sotto controllo;
  - le specifiche del prodotto sono molto vincolate o la produzione è particolarmente delicata;
  - l'operatore deve decidere, se modificare il processo o come valutare un certo settaggio;
  - si richiede una modifica nelle specifiche del prodotto;
  - deve essere continuamente certificata la capacità del processo.
- Le carte per attributi  $p, c, u$  devono essere usate per valutare un prodotto nei seguenti casi:
  - è richiesta una riduzione del numero di pezzi non funzionanti;
  - il prodotto è così complesso che l'unico strumento di verifica è l'attributo di funzionamento o guasto;
  - non è possibile effettuare misure delle grandezze osservate;
  - è richiesta la storia passata della produzione.

Le carte per attributi sono estremamente sintetiche e particolarmente utili, per effettuare comparazioni tra processi.

- Le carte per misure singole, usate insieme alle carte con range mobile, sono utili nei seguenti casi:
  - non è possibile disporre più di un dato, per campione, o la replica dell'esperimento non dipende dal laboratorio che la effettua;
  - sono disponibili strumenti di controllo automatici che effettuano il controllo di tutte le unità;
  - i dati sono disponibili solo con grande lentezza, al punto da rendere impraticabile la realizzazione di un campione con più di una unità.

#### Azioni da intraprendere per migliorare il processo

L'applicazione delle carte di controllo permette di avere informazioni, su due aspetti salienti del processo:

- il suo stato di controllo;

- la sua capacità.

La tabella sottostante dà risposta ai quesiti: se il processo è sotto controllo e se il processo ha capacità sufficiente.

			Il processo ha sufficiente capacità?	
			Si	No
Il processo è sotto controllo?	Si	SPC	SPC Piani sperimentali Verifica delle specifiche del processo Modifica del processo	
	No	SPC	SPC Piani sperimentali Verifica delle specifiche del processo Modifica del processo	

A seconda delle caselle, la tabella propone diversi tipi di interventi.

- La situazione ideale è nella casella in alto a sinistra, in corrispondenza della quale va praticamente tutto bene ed i metodi di SPC sono richiesti per garantire il mantenimento di un certo standard qualitativo.
- La casella in alto a destra evidenzia invece una situazione in cui, a fronte di uno stato di controllo, si assiste alla produzione di molte unità difettose, probabilmente per l'eccessiva variabilità del processo o per la scelta dei limiti di specifica troppo stretti, per la cui revisione si può procedere con un'indagine sperimentale e, al limite, con la sostituzione del processo con uno più moderno.
- La casella in basso a destra evidenzia la situazione peggiore per la quale i metodi di SPC devono fornire velocemente indicazioni sulla direzione in cui agire, per migliorare la qualità.
- La situazione rappresentata dalla casella in basso a sinistra corrisponde infine al caso in cui la scelta dei limiti di specifica sono troppo ampi e, in questo caso, i metodi di SPC devono aiutare la loro revisione, a beneficio di una migliore immagine dell'efficacia del processo.

#### Strumenti per la raccolta dei dati per l'analisi di SPC

Negli ultimi anni, si è assistito ad un considerevole aumento dell'offerta software per SPC. Ad esempio, la rivista *Quality Progress* riporta un elenco di software disponibili, suddivisi per categorie, e le corrispondenti società produttrici. In questa attività, l'uso del computer è particolarmente importante. Infatti:

- inizialmente l'applicazione dei metodi di SPC deve essere fatta manualmente, per apprendere meglio gli strumenti, ma in seguito le operazioni richieste, per eseguire i calcoli, possono diventare troppo lunghe;
- l'abitudine a memorizzare i dati, in formato standard, permette di scambiare velocemente informazioni, tra diversi settori;
- la presenza di un sistema informatico permette altresì di effettuare controlli, anche su più variabili, in brevissimo tempo, aumentando la produttività dell'operatore.

## 7. Analisi di capacità del processo

Il concetto di capacità del processo, già formalmente introdotto, confronta la variabilità naturale del processo con le specifiche del prodotto. Tra i vari strumenti dedicati a questo scopo, la carta di controllo costituisce la tecnica, più semplice ed efficace, di analisi della capacità del processo.

### 7.1 Introduzione

L'analisi di capacità del processo, in generale, è riferita all'uniformità di comportamento del processo. In questo contesto, le tecniche statistiche sono utili durante tutto il ciclo di produzione, a partire dalle attività di sviluppo, preliminari alla fabbricazione, per continuare con l'analisi e la valutazione della variabilità del processo, in relazione ai livelli nominali di specifica, fino alle operazioni dedicate all'eliminazione od almeno alla riduzione di questa variabilità che è una misura dell'uniformità della caratteristica del prodotto in uscita, interpretabile come:

- variabilità naturale od inerente ad uno specifico istante, detta anche variabilità istantanea;
- variabilità rispetto al tempo.

E' consuetudine assumere l'intervallo di ampiezza 6-sigma, della distribuzione della caratteristica di qualità del prodotto, come misura della capacità del processo e, in un processo la cui caratteristica di qualità ha distribuzione normale (con media  $\mu$  e deviazione standard  $\sigma$ ), i limiti di tolleranza naturale, superiore ed inferiore (UNTL e LNTL), posti ai valori:  $\mu \pm 3\sigma$ , sono:

$$UNTL = \mu + 3\sigma$$

$$LNTL = \mu - 3\sigma$$

Per la distribuzione normale, l'intervallo tra i limiti di tolleranza naturale corrisponde ad una probabilità del 99.73%, avendo solo una probabilità dello 0.27% di ottenere valori fuori da detto intervallo. A riguardo:

- anche se una probabilità pari a 0.27% può sembrare piccola, essa corrisponde comunque a 2700 unità non conformi, ogni milione;
- se la distribuzione della variabile in uscita del processo non è gaussiana, la percentuale di elementi fuori dai limiti:  $\mu \pm 3\sigma$ , può anche risultare considerevolmente diversa da 0.27%.

L'analisi di capacità del processo è la procedura di stima della capacità, effettuata con riferimento alla forma della distribuzione di probabilità, alla sua media e alla sua deviazione standard. Ad esempio, nel caso di una distribuzione normale (con una data media ed una data deviazione standard), si può condurre l'analisi di capacità, senza riferimento alle specifiche di qualità del prodotto. In alternativa, è invece possibile esprimere la capacità del processo come percentuale di elementi fuori specifica.

Uno studio di capacità di un processo misura spesso alcuni parametri funzionali del processo, non l'intero processo. Se un analista potesse misurare direttamente il processo e controllare direttamente l'attività di rilevazione dei dati, lo studio sarebbe un vero studio di capacità, perché potrebbe effettivamente realizzare una procedura inferenziale, indirizzata allo studio della stabilità del processo.

Invece avendo a disposizione solo le unità campionarie, fornite dal fornitore, senza poter osservare direttamente il processo, né possedere la conoscenza storica del processo di produzione, lo studio dovrebbe, più propriamente, essere chiamato di caratterizzazione del prodotto. Infatti in questo tipo di studio, è analizzata la distribuzione della caratteristica di qualità (frazione di elementi conformi) e non il comportamento dinamico del processo, eventualmente osservato nel solo stato di controllo statistico.

L'analisi di capacità del processo è parte vitale di un programma complessivo di miglioramento della qualità e prevede di utilizzare i dati, principalmente allo scopo di:

- prevedere come il processo rispetti le tolleranze;
- assistere i tecnici del servizio, di ricerca e sviluppo, nelle modifiche di processo;
- stabilire l'intervallo di campionamento, per le procedure di sorveglianza;
- stabilire i requisiti di prestazione di nuove attrezzature;
- selezionare i fornitori;
- pianificare la produzione, anche in presenza di interazione del processo sulle tolleranze;
- ridurre la variabilità del processo.

In questo modo, l'analisi della capacità del processo ha applicazione in più parti del ciclo di vita di un prodotto, inclusa la fase di progettazione, quella di scelta dei fornitori, di pianificazione della produzione e quella dell'effettiva realizzazione del prodotto. A riguardo, per l'analisi della capacità del processo sono usate principalmente tre tecniche: istogrammi o carte di probabilità, carte di controllo e programmazione degli esperimenti.

## 8. Altre tecniche statistiche di monitoraggio e controllo del processo

I metodi di base dell'analisi statistica di capacità e di controllo del processo fanno uso di molte tecniche, come le carte di controllo di Shewhart (utilizzate per oltre 50 anni). Ad ogni modo, la crescente attenzione alla riduzione della variabilità ed al miglioramento dei processi, insieme al successo dei metodi di base, ha portato allo sviluppo di molte nuove tecniche per il monitoraggio ed il controllo statistico del processo.

### 8.1 Carte di controllo CUSUM ed EWMA

La caratteristica principale delle carte (di controllo) di Shewhart è l'uso esclusivo delle informazioni sul processo, contenute nell'ultimo istante di osservazione, ignorando tutte le precedenti. Tutto questo rende la carta di Shewhart relativamente insensibile a piccole modifiche del livello del processo, di ampiezza in genere non superiore ad  $1.5\sigma$ . Per questo motivo, è suggerito di corredare queste carte con ulteriori strumenti, quali il test delle sequenze, oppure l'uso dei limiti di preallarme, allo scopo di considerare, nella procedura decisionale, tutte le informazioni disponibili. Tuttavia queste procedure, miranti a rendere più sensibile la carta, ne riducono la semplicità d'uso e l'immediatezza interpretativa.

Due valide alternative alla carta di controllo di Shewhart, in presenza di limitati salti di livello, sono:

- le carte **CUSUM** (*CUmulative SUM*) od **a somme cumulate**:

$$C_i^+ = \max(0, C_{i-1}^+ + y_i - K)$$

$$C_i^- = \max(0, C_{i-1}^- - y_i - K)$$

dati:  $y_i = x_i - \mu_0$  oppure  $y_i = (x_i - \mu_0)/\sigma$  (per le carte standardizzate)

fissati:  $C_0^+ = C_0^- = 0$  valori iniziali

inoltre  $K = |\mu_1 - \mu_0|/2$  valore di tolleranza (essendo  $\mu_1$  il possibile outlier cercato)

$H = 5\sigma$  limite di tolleranza (per il fuori controllo:  $C_i^+ > H$  or  $C_i^- > H$ )

- le carte **EWMA** (*Exp. Weighted Moving Average*) ovvero **a medie mobili pesate esponenzialmente** ed **EWMS** (*Exp. Weighted Mean Squared*) ovvero **per la variabilità pesata esponenzialmente**:

$$UCL_i = \mu + L\sigma\sqrt{\lambda/(2-\lambda) \cdot (1-(1-\lambda)^{2i})}$$

$$UCL = \sigma\sqrt{\chi_{v,\alpha/2}^2/v}$$

$$CL = \mu$$

$$CL = \sigma$$

$$LCL_i = \mu - L\sigma\sqrt{\lambda/(2-\lambda) \cdot (1-(1-\lambda)^{2i})}$$

$$LCL = \sigma\sqrt{\chi_{v,1-\alpha/2}^2/v}$$

dati:  $y_i = x_i + (1-\lambda)y_{i-1}$

$$S_i^2 = (1-\lambda)S_{i-1}^2 + \lambda(x_i - \mu)^2$$

fissati:  $\lambda = 0.05 \div 0.2$

da cui  $S_i = \sqrt{S_i^2}$

inoltre  $L = 3\sigma$

(errore quadratico medio ponderato esponenzialmente)

### 8.1.1 La procedura maschera a V

La **maschera a V** è applicata ai valori successivi della statistica cumulata <sup>3</sup>:

$$C_i = \sum_{j=1}^i y_j = C_{i-1} + y_i$$

dove  $y_i = (x_i - \mu)/\sigma$  è la generica osservazione standardizzata. La procedura decisionale consiste nel posizionare la maschera a V in un punto del grafico, in corrispondenza dell'ultimo valore  $C_i$  disponibile; in questo modo:

- se i punti corrispondenti alle precedenti somme cumulate  $C_1, C_2, \dots, C_i$  giacciono tutti all'interno dei due bracci della maschera a V, il processo è da considerare in stato di controllo.
- se invece almeno uno dei punti è fuori dalla maschera, il processo è da considerare fuori controllo.

<sup>3</sup> Nelle utilizzazioni pratiche, la maschera deve essere applicata orizzontalmente, ad ogni punto del grafico della carta CUSUM, ed i suoi bracci tracciati all'indietro, in direzione dell'origine. Le prestazioni della maschera a V dipendono dalla distanza della linea verticale di taglio (della zona di accettazione), dal vertice della V, e dall'angolo di apertura dei bracci. Comunque la maschera a V presenta alcuni svantaggi e problemi:

- la maschera a V è uno schema di controllo bilaterale, inutilizzabile nel caso di problemi unilaterali;
- la determinazione dell'estensione assegnata ai bracci della maschera a V è talvolta difficoltosa;
- la maschera a V è fortemante sensibile all'ambiguità associata ai valori del livello di significatività e della potenza del test.

## 9. Progetto e miglioramento del processo tramite la programmazione statistica degli esperimenti

I miglioramenti della qualità sono più efficaci, quando sono parte integrante del ciclo di sviluppo del prodotto e del processo. In particolare, l'introduzione formale di una metodologia di programmazione statistica degli esperimenti, nei primi stadi del ciclo di sviluppo, è spesso la chiave del successo complessivo, laddove si tratta di progettazione di nuovi prodotti, di miglioramento del progetto di prodotti esistenti e di ottimizzazione del processo. Questo principio è verificato in molti settori, perché l'uso efficace di una solida metodologia di programmazione statistica degli esperimenti può portare a prodotti, più facili da realizzare, con una migliore affidabilità e migliori prestazioni sul campo. La programmazione statistica degli esperimenti può anche migliorare molto le attività di sviluppo del processo di messa a punto.

In questo contesto, l'analisi di varianza è usata come base per analizzare i dati provenienti da esperimenti programmati. Infatti introdurre la programmazione degli esperimenti, senza fare ricorso a questo metodo, è un errore, soprattutto perché gli utenti incontrano comunque l'analisi di varianza, in ogni software utilizzato.

### 9.1 I principi di base della programmazione statistica degli esperimenti

La terminologia della programmazione statistica degli esperimenti è introdotta con la presentazione di alcune brevi illustrazioni dei tipi di problemi in cui la programmazione degli esperimenti è utile, discutendo i principali metodi statistici usati. Un obiettivo importante sviluppa una comprensione dell'analisi di varianza, tecnica usata per la maggior parte degli esperimenti programmati. Altri argomenti chiave comprendono il ruolo del modello, sottostante l'esperimento, i residui ed il loro uso, nella verifica delle ipotesi ed il principio del bloccaggio, utile negli esperimenti con variabili di disturbo coinvolte.

Un esperimento programmato è una prova od una serie di prove, dove sono fatte variare deliberatamente le variabili di ingresso di un processo, in modo da poter osservare ed identificare le corrispondenti variazioni della risposta in uscita. Infatti un processo può essere visualizzato, come un complesso di macchine, metodi e persone che trasformano il materiale in ingresso nel prodotto in uscita. Questo prodotto in uscita ha una o più caratteristiche di qualità o risposte osservabili. Alcune variabili di processo sono controllabili, mentre altre sono incontrollabili (anche se queste ultime possono diventare controllabili, nel corso dell'esperimento, ed a volte, questi fattori incontrollabili sono detti fattori di rumore). Gli obiettivi dell'esperimento comprendono:

- la determinazione di quali variabili hanno più influenza sulla risposta;
- la determinazione di come aggiustare le variabili più influenti, in modo che la risposta sia vicina alla richiesta della specifica;
- la determinazione di come aggiustare le variabili influenti, in modo che la variabilità della risposta sia piccola;
- la determinazione di come aggiustare le variabili influenti, in modo che l'effetto delle altre variabili non controllabili sia minimizzato.

I metodi della programmazione degli esperimenti possono essere usati nello sviluppo e nella messa a punto del processo, per migliorare le prestazioni di un processo od ottenere un processo robusto, cioè insensibile alle sorgenti esterne di variabilità.

I metodi di controllo statistico del processo e la programmazione statistica degli esperimenti, due strumenti potenti per il miglioramento e l'ottimizzazione del processo, sono strettamente collegati. Ad esempio, se un

processo è sotto controllo statistico, ma ha ancora una bassa capacità, è necessario ridurre la variabilità per migliorare il processo. Esperimenti programmati possono fornire un modo di ottenere il risultato più efficiente del semplice controllo statistico, del processo SPC. In sostanza, l'SPC è un metodo statistico passivo, dove si osserva il processo, in attesa di certe informazioni che possono condurre a cambiamenti utili. Tuttavia se il processo è sotto controllo, l'osservazione passiva non può produrre molta informazione utile. Viceversa la programmazione statistica degli esperimenti è un metodo statistico attivo, dove si fanno, in effetti, una serie di prove sul processo, effettuando cambiamenti negli ingressi ed osservando i corrispondenti cambiamenti nelle uscite, cosa che produce informazione e può condurre ad un miglioramento del processo.

## **9.2 Vantaggi nel miglioramento della qualità e del processo**

I metodi di programmazione statistica degli esperimenti possono anche essere molto utili nella messa sotto controllo statistico del processo. Ad esempio, si supponga che una carta di controllo indichi un processo fuori controllo, ma con molte variabili di ingresso controllabili. D'altra parte, può essere difficile ricondurre il processo sotto controllo a meno che si non conoscano quali variabili di ingresso hanno importanza. I metodi di programmazione statistica degli esperimenti possono essere usati, per identificare le variabili di ingresso più influenti.

La programmazione statistica degli esperimenti è uno strumento di importanza critica, per il miglioramento di un processo, e ha anche estese applicazioni nello sviluppo di nuovi processi, in quanto l'applicazione tempestiva di queste tecniche, nello sviluppo di processi, può condurre a:

- un aumento del volume prodotto;
- una riduzione della variabilità ed un più preciso rispetto delle specifiche;
- una riduzione dei tempi di sviluppo;
- una riduzione dei costi complessivi.

I metodi di programmazione degli esperimenti possono anche giocare un ruolo importante nelle attività di progetto, quando si sviluppano nuovi prodotti o si migliorano quelli esistenti. Infatti alcune applicazioni della programmazione statistica degli esperimenti comprendono:

- la valutazione ed il confronto di fondamentali configurazioni di progetto;
- la valutazione di alternative sui materiali;
- la determinazione dei parametri di progetto chiave, in relazione alla loro influenza sulle prestazioni.

L'uso della programmazione statistica degli esperimenti, in queste aree, può condurre ad un miglioramento della fabbricabilità del prodotto, a migliorate prestazioni operative e di affidabilità, a tempi di sviluppo del prodotto più corti ed a minori costi di produzione.

## **9.3 Criteri per la programmazione degli esperimenti**

Gli esperimenti programmati costituiscono un approccio potente al miglioramento del processo. Tuttavia per poter usare questo approccio, è necessario che tutti i coinvolti nell'esperimento abbiano preliminarmente una chiara idea del suo obiettivo, un'idea esatta dei fattori da studiare, una conoscenza di come l'esperimento

deve essere eseguito ed una comprensione, almeno qualitativa, di come i dati sono analizzati, secondo una determinata procedura:

- riconoscimento ed enunciazione del problema;
- scelta dei fattori e dei loro livelli;
- selezione della variabile di risposta;
- scelta del piano sperimentale;
- esecuzione dell'esperimento;
- analisi dei dati;
- conclusioni e raccomandazioni.

Nel corso dell'intero processo, è importante sapere che la sperimentazione è una parte del processo, di conoscenza, nella quale si formulano ipotesi di lavoro, riguardo al sistema, poi si eseguono esperimenti, per investigare queste ipotesi, e si formulano nuove ipotesi, sulla base dei risultati, e così via. Questo suggerisce che la sperimentazione è iterativa.

Normalmente è un grave errore programmare, all'inizio dello studio, un unico, grande ed esaustivo, esperimento. Un esperimento, di successo, richiede la conoscenza di quali sono i fattori importanti, dell'intervallo in cui questi fattori sono poi fatti variare, dell'appropriato numero di livelli da usare e delle appropriate unità di misura, per tutte queste variabili. In generale, non si conosce la risposta a queste domande, ma si impara qualcosa, su esse, nel corso del cammino.

Man mano che un programma di esperimenti procede, spesso si lasciano cadere alcune variabili, se ne aggiungono altre, si cambia la regione di esplorazione di alcuni fattori, si aggiungono anche nuove variabili di risposta. In conclusione, di solito si sperimenta in modo sequenziale e, come regola generale, non più di un quarto delle risorse disponibili deve essere investito nel primo esperimento. Tutto questo assicura risorse disponibili sufficienti, per raggiungere l'obiettivo finale dell'esperimento.

#### 9.4 Esperimenti fattoriali

I piani fattoriali devono essere usati ogni volta che, in un esperimento, si hanno molti fattori da studiare. In questo tipo di piano, tutti i fattori sono fatti variare insieme. Precisamente con l'espressione "esperimento fattoriale" si intende che, in ogni insieme completo di prove od in ogni sua replicazione, sono studiate tutte le possibili combinazioni di livelli dei fattori e, se si hanno due fattori  $A$  e  $B$ , rispettivamente con  $a$  e  $b$  livelli, allora ogni replicazione contiene tutte le  $ab$  possibili combinazioni.

Si definisce effetto del fattore la variazione nella risposta prodotta dal cambiamento di livello del fattore. Questo si chiama effetto principale, perché si riferisce ai fattori principali dello studio. Si considerino ad esempio i dati della Figura 9.4.1.

In questo piano fattoriale entrambi i fattori  $A$  e  $B$  hanno due livelli, indicati rispettivamente con  $-$  e  $+$ , e detti rispettivamente: "basso" e "alto". L'effetto principale, del fattore  $A$ , è la differenza, fra la risposta media, al livello alto di  $A$ , e la risposta media, al livello basso di  $A$ :

$$A = \bar{y}_A - \bar{y}_A = \frac{30 + 40}{2} - \frac{10 + 20}{2} = 20$$

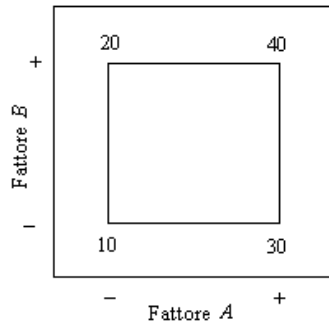


Fig. 9.4.1 – Esperimento fattoriale con due fattori

In altre parole, il passaggio del fattore  $A$  dal livello basso  $-$  al livello alto  $+$  causa in media un aumento della risposta di 20 unità. Analogamente, l'effetto principale di  $B$  è:

$$B = \bar{y}_B - \bar{y}_B = \frac{20 + 40}{2} - \frac{10 + 30}{2} = 10$$

In alcuni esperimenti, si ha un'interazione tra i fattori e la differenza di risposta fra i livelli di un fattore non è la stessa a tutti i livelli degli altri fattori. Ad esempio, data la Figura 9.4.2, ai livelli basso ed alto del fattore  $B$  l'effetto di  $A$  sono rispettivamente:

$$A = 30 - 10 = 20 \qquad A = 0 - 20 = -20$$

cosicché si ha interazione fra  $A$  e  $B$ , dato che l'effetto di  $A$  dipende dal livello scelto per il fattore  $B$ .

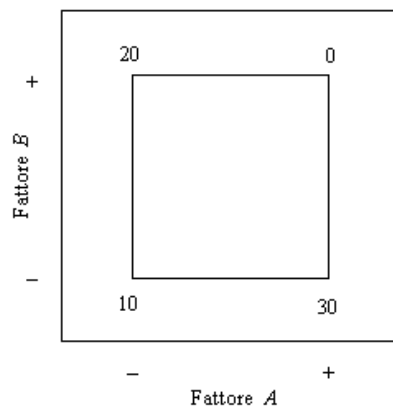


Fig. 9.4.2 – Esperimento fattoriale con due fattori ed interazione

Quando l'interazione è grande, i corrispondenti effetti principali sono poco significativi. Ad esempio, se si usano i dati della Figura 9.4.2, l'effetto principale di  $A$  vale:

$$A = \frac{30 + 0}{2} - \frac{10 + 20}{2} = 0$$

e si potrebbe concludere che non esista l'effetto principale di  $A$ . D'altra parte, se si esamina l'effetto principale di  $A$ , a livelli diversi del fattore  $B$ , le cose non stanno così, perché l'effetto del fattore principale di  $A$ , dipende dai livelli del fattore  $B$ . Allora l'interazione  $AB$  è maggiore dell'effetto principale di  $A$  e pertanto un'interazione significativa può mascherare la significatività degli effetti principali. Il concetto di interazione può essere illustrato graficamente, come mostrano le Figure 9.4.3 e 9.4.4.

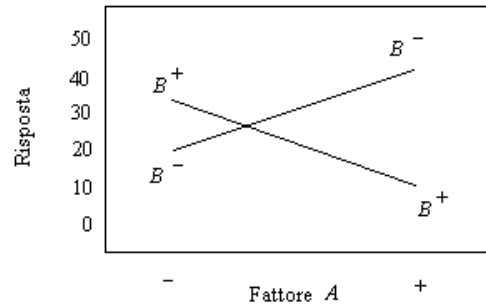
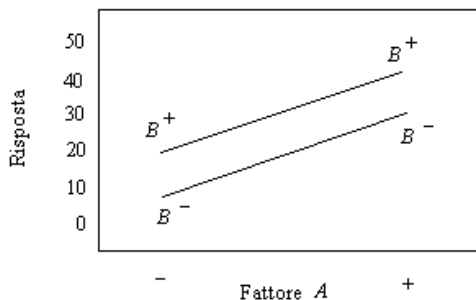


Fig. 9.4.3 – Esperimento fattoriale senza interazione      Fig. 9.4.4 – Esperimento fattoriale con interazione

La Figura 9.4.3 contiene una rappresentazione dei dati della Figura 9.4.1, in funzione dei livelli di  $A$  per entrambi i livelli di  $B$ , dove le due linee  $B^-$  e  $B^+$  sono approssimativamente parallele ed i fattori  $A$  e  $B$  non interagiscono.

Invece la Figura 9.4.4 rappresenta i dati della Figura 9.4.2; in questo grafico, le linee  $B^-$  e  $B^+$  non sono parallele, indicando una interazione fra i fattori  $A$  e  $B$ .

## 9.5 Piani fattoriali $2^k$

Alcuni tipi particolari di piani fattoriali sono molto utili nello sviluppo e nel miglioramento del processo. Uno di questi è il piano fattoriale con  $k$  fattori, ciascuno con due livelli. Dato che ogni replicazione completa del piano ha  $2^k$  prove, questa disposizione è detta piano fattoriale  $2^k$ .

### 9.5.1 Il piano $2^2$

Il più semplice piano  $2^k$  è il piano  $2^2$ , composta da due fattori  $A$  e  $B$ , ciascuno con due livelli, indicati rispettivamente con  $-$  e  $+$ , e detti rispettivamente: "basso" e "alto". Il piano  $2^2$  è già illustrato nelle figure precedenti e può essere rappresentato geometricamente come un quadrato, con  $2^2 = 4$  prove a formare i vertici del quadrato.

In generale, per rappresentare le prove si usa una notazione speciale, dove una prova è rappresentata da una serie di lettere minuscole e, se una lettera è presente, il corrispondente fattore è fissato al valore alto, mentre se è assente, il fattore è tenuto al livello basso.

Ad esempio, la prova  $a$  indica che il fattore  $A$  è tenuto a livello alto, mentre il fattore  $B$  è a livello basso. La prova con entrambi i valori a livello basso è indicata con  $(1)$ . Questa notazione è usata per tutta la serie di piani  $2^k$ . Ad esempio, la prova in un piano  $2^4$ , dove i fattori  $A$  e  $C$  sono a livello alto, mentre i fattori  $B$  e  $D$  sono a livello basso, è indicata con  $ac$ .

Gli effetti di interesse, in un piano  $2^2$ , sono i due effetti principali  $A$  e  $B$  e l'interazione fra i due fattori  $AB$ . Indicando con i simboli:  $(1)$ ,  $a$ ,  $b$ ,  $ab$ , i totali di tutte le  $n$  osservazioni in questi punti del piano, è facile stimare gli effetti di questi fattori.

- Per stimare l'effetto principale di  $A$ , occorre prendere la media delle osservazioni sul lato destro del quadrato, dove  $A$  è a livello alto, e da questa sottrarre la media delle osservazioni sul lato sinistro del quadrato, dove  $A$  è a livello basso:

$$A = \bar{y}_{A^+} - \bar{y}_{A^-} = \frac{a + ab}{2n} - \frac{b + (1)}{2n} = \frac{a + ab - b - (1)}{2n}$$

- Analogamente l'effetto principale di  $B$  si determina facendo la media delle osservazioni sulla parte superiore del quadrato, dove  $B$  è al livello alto, e da questa sottraendo la media delle osservazioni sulla parte inferiore del quadrato, dove  $B$  è al livello basso:

$$B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{b + ab}{2n} - \frac{a + (1)}{2n} = \frac{b + ab - a - (1)}{2n}$$

- Infine, l'interazione  $AB$  si stima prendendo la differenza delle medie sulle diagonali del quadrato, corrispondente piano  $2^2$ :

$$AB = \frac{ab + (1)}{2n} - \frac{a + b}{2n} = \frac{ab + (1) - a - b}{2n}$$

I numeratori delle suddette equazioni sono detti contrasti (ad esempio, il contrasto  $A$  è:  $a + ab - b - (1)$ ) e, in tutte queste equazioni i coefficienti dei contrasti sono sempre  $+1$  o  $-1$ . Per determinare i segni in ogni prova di un particolare contrasto, si può usare la seguente tabella di segni.

		<i>Effetti fattoriali</i>				
		$I$	$A$	$B$	$AB$	
	1	$(1)$	+	-	-	+
<i>Combinazioni di trattamenti</i>	2	$a$	+	+	-	-
	3	$b$	+	-	+	-
	4	$ab$	+	+	+	+

Le intestazioni delle colonne sono gli effetti principali di  $A$  e di  $B$ , l'effetto dell'interazione  $AB$  e quello del totale  $I$ . Le intestazioni delle righe sono le prove. Per generare i contrasti, a partire da questa tabella, si moltiplicano i segni della colonna appropriata per la combinazione di livelli di fattori, elencati nella seconda colonna, e poi si sommano tutti gli addendi (a riguardo, si osservi che i segni della colonna  $AB$  sono i prodotti dei segni delle colonne  $A$  e  $B$ ).

Usando il seguente risultato:

$$SS = \frac{(\text{contrasto})^2}{n \sum (\text{coefficienti dei contrasti})^2}$$

si ottengono le somme dei quadrati per  $A$ ,  $B$  ed  $AB$ :

$$SS_A = \frac{(a + ab - b - (1))^2}{4n} \quad SS_B = \frac{(b + ab - a - (1))^2}{4n} \quad SS_{AB} = \frac{(ab + (1) - a - b)^2}{4n}$$

L'analisi della varianza è completata calcolando la somma dei quadrati  $SS_{totale}$  (con  $4n - 1$  gradi di libertà) e ricavando la somma dei quadrati dell'errore  $SS_{errore}$  (con  $4(n - 1)$  gradi di libertà), per sottrazione.

#### Un esempio: l'esperimento sulla fresatrice

Una fresatrice verticale è utilizzata per tagliare tacche di registrazione su basette di un circuito stampato. La media delle dimensioni della tacca è soddisfacente ed il processo è sottocontrollo statistico (dalle carte di controllo  $\bar{x}$  e  $R$ ), ma si ha troppa variabilità nel processo. Questo eccesso di variabilità produce problemi nell'assemblaggio della basetta. Infatti i componenti sono inseriti, usando l'apparecchiatura automatica, e la variabilità nella dimensione della tacca produce uno scorretto posizionamento della basetta, cosicché, il dispositivo di inserimento automatico non funziona correttamente.

Dato che il processo è in stato di controllo statistico, il gruppo per il miglioramento della qualità del progetto decide di usare un esperimento programmato, per studiare il processo, con un piano  $2^2$ , prendendo in considerazione due fattori: la dimensione della punta da taglio  $A$  e la velocità  $B$ .

Per ciascuno dei due fattori, sono scelti due livelli, la dimensione della punta  $A$  a  $1/16''$  ed a  $1/8''$ , e la velocità  $B$  a 40 ed a 80 giri al minuto ( $rpm$ ). Dato che la variazione nella dimensione della tacca risulta difficile da misurare direttamente, il gruppo decide di misurarla indirettamente. A tal fine, sono attrezzate sedici basette, con accelerometri che permettono la misurazione delle vibrazioni nei tre assi coordinati. Il vettore risultante di queste tre misurazioni è utilizzato come variabile di risposta. Dato che la vibrazione della superficie della basetta durante il taglio è direttamente legata alla variabilità della dimensione della tacca, una riduzione del livello di vibrazione riduce anche la variabilità nelle dimensioni della tacca.

Allora testate quattro piastre, per ciascuna delle quattro prove dell'esperimento, i risultati ottenuti sono dati nella seguente tabella.

Prova		Fattori		Vibrazioni				Totale
		A	B					
1	(1)	-	-	18.2	27.2	15.9	41.0	64.4
2	a	+	-	18.9	24.0	14.5	43.9	96.1
3	b	-	+	+12.9	22.4	15.1	36.3	59.7
4	ab	+	+	14.4	22.5	14.2	39.9	161.7

Pertanto usando le sopraccitate equazioni del piano  $2^2$ , si possono stimare gli effetti dei vari fattori presenti:

$$A = \frac{a + ab - b - (1)}{2n} = \frac{133.1}{8} = 16.64 \qquad B = \frac{b + ab - a - (1)}{2n} = \frac{60.3}{8} = 7.54$$

$$AB = \frac{ab + (1) - a - b}{2n} = 8.71$$

Tutti i valori numerici delle stime degli effetti appaiono grandi. Ad esempio, facendo passare il fattore  $A$  dal valore basso a quello alto (cioè dimensione della punta da  $1/16''$  ad  $1/8''$ ), il livello medio di vibrazione aumenta di  $16.64 \text{ cps}$ . La grandezza di questi effetti è confermata dall'analisi della varianza, riassunta nella seguente tabella.

<i>Sorgente di variazione</i>	<i>Somma dei quadrati</i>	<i>Gradi di libertà</i>	<i>Media dei quadrati</i>	$F_0$	$P$
$A$	1107.226	1	1107.226	185.25	$1.17 \times 10^{-8}$
$B$	227.256	1	227.256	38.03	$4.82 \times 10^{-5}$
$AB$	303.631	1	303.631	50.80	$1.20 \times 10^{-5}$
<i>Errore</i>	71.723	12	5.977		
<i>Totale</i>	1709.836	15			

L'analisi della varianza conferma le conclusioni, ottenute inizialmente dalla osservazione della grandezza e della direzione degli effetti dei fattori; infatti sia la dimensione della punta da taglio, sia la velocità sono importanti ed è presente anche una interazione fra le due variabili.

#### Modello di regressione ed analisi dei residui

È facile ottenere i residui da un piano  $2^2$ , adattando un modello di regressione ai dati. Per l'esperimento della fresatrice, il modello di regressione è costituito dalla relazione:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

dove i fattori  $A$  e  $B$  sono rappresentati dalle variabili codificate  $x_1$  e  $x_2$ , e l'interazione  $AB$  è rappresentata dal termine di prodotto misto  $x_1 x_2$  del modello.

Ai livelli basso e alto, di ciascun fattore, sono assegnati rispettivamente i valori:  $x_j = -1$  e  $x_j = 1$ . I coefficienti  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  e  $\beta_3$  sono detti coefficienti di regressione ed  $\varepsilon$  è un termine di errore casuale, analogo al termine di errore, nel modello di analisi della varianza, cosicché il modello di regressione adattato risulta:

$$\hat{y} = 23.83 + \frac{16.64}{2} x_1 + \frac{7.54}{2} x_2 + \frac{8.71}{2} x_1 x_2$$

In questa espressione, la stima della intercetta  $\beta_0$  è la media generale di tutte le 16 osservazioni  $\bar{y}$  e la stima degli altri coefficienti di regressione  $\beta_j$  è la metà della stima dell'effetto del corrispondente fattore (infatti ogni stima di coefficiente di regressione è metà della stima dell'effetto, perché esso misura l'effetto di una variazione unitaria di  $x_j$  sulla media di  $y$  e la stima dell'effetto è basata su una variazione su due unità: da  $-1$  a  $+1$ ).

Questo modello può essere utilizzato per ottenere i valori stimati del livello di vibrazione, in ogni punto della regione di sperimentazione, inclusi i quattro punti del piano. Ad esempio, dato il punto con piccola dimensione della punta:  $x_1 = -1$ , e con bassa velocità:  $x_2 = -1$ , il livello stimato di vibrazione è pari a:

$$\hat{y} = 23.83 + \frac{16.64}{2}(-1) + \frac{7.54}{2}(-1) + \frac{8.71}{2}(-1)(-1) = 16.1$$

ed i quattro residui, corrispondenti alle osservazioni in questo punto del piano, si trovano prendendo le differenze fra le osservazioni effettive ed il valore previsto:

$$\begin{aligned} e_1 &= 18.2 - 16.1 = 2.1 & e_2 &= 18.9 - 16.1 = 2.8 \\ e_3 &= 12.9 - 16.1 = -3.2 & e_4 &= 14.4 - 16.1 = -1.7 \end{aligned}$$

I livelli stimati di vibrazione ed i residui, negli altri tre punti del piano, sono calcolati in modo analogo.

#### Procedura di analisi per gli esperimenti fattoriali

La tabella riassume la successione dei passi, usualmente eseguiti nell'analisi degli esperimenti fattoriali.

- stimare gli effetti dei fattori;
- costruire il modello iniziale;
- testare la significatività degli effetti dei fattori;
- analizzare i residui;
- migliorare il modello;
- interpretare i risultati.

Il modello preliminare, usato nell'analisi, è il modello fattoriale a due fattori, con interazione e, in generale, in ogni esperimento fattoriale, con replicazioni, si usa quasi sempre il modello fattoriale completo, come modello iniziale.

#### **9.5.2 Il piano $2^k$ con $k \geq 3$**

I metodi per i piani fattoriali, con  $k = 2$  fattori, possono essere facilmente estesi, a tutti i casi, con più di due fattori.

Ad esempio, considerato il caso  $k = 3$  fattori, ciascuno a due livelli, questo piano è un piano fattoriale  $2^3$  e presenta otto combinazioni di livelli dei fattori. Geometricamente è rappresentato da un cubo (come mostra la Figura 9.5.1), con le otto prove a costituire i vertici del cubo.

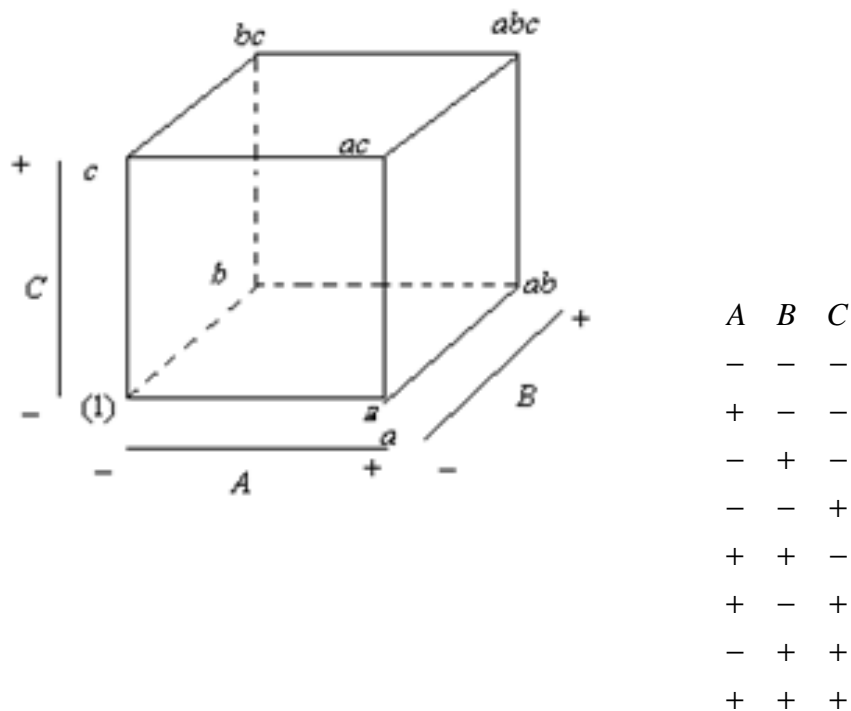


Fig. 9.5.1 – Esperimento fattoriale con tre fattori

Questo piano permette di stimare tre effetti principali ( $A$ ,  $B$  e  $C$ ), unitamente a tre interazioni fra due fattori ( $AB$ ,  $AC$  e  $BC$ ) e con una interazione tra i tre fattori ( $ABC$ ), dove il modello fattoriale completo è scritto simbolicamente come:

$$y = \mu + A + B + C + AB + AC + BC + ABC + \varepsilon$$

dove  $\mu$  è la media generale,  $\varepsilon$  un termine di errore casuale, supposto distribuito come  $N(0, \sigma^2)$  indipendenti, e le lettere maiuscole rappresentano gli effetti principali e le interazioni.

Invece le lettere minuscole (1),  $a$ ,  $b$ ,  $c$ ,  $ab$ ,  $ac$ ,  $bc$  ed  $abc$  rappresentano il totale di tutte le  $n$  replicazioni, per ciascuna delle otto prove del piano. Anche in questo caso, gli effetti principali possono essere stimati facilmente e, riferendosi al cubo, per stimare l'effetto principale di  $A$ , si deve fare la media delle quattro prove sulla faccia destra del cubo, dove  $A$  è al livello alto, sottraendo da questa quantità la media delle quattro prove sulla faccia sinistra del cubo, dove  $A$  è al livello basso:

$$A = \bar{y}_{A^+} - \bar{y}_{A^-} = \frac{a + ab + ac + abc - b - c - bc - (1)}{4n}$$

In modo analogo, l'effetto principale di  $B$  è la differenza fra la media delle quattro prove sulla faccia posteriore del cubo meno quella delle quattro prove sulla faccia di fronte:

$$B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{b + ab + bc + abc - a - c - ac - (1)}{4n}$$

ed ancora l'effetto principale di  $C$  è la differenza fra la media delle quattro prove sulla faccia in alto del cubo meno quella delle quattro prove sulla faccia in basso.

$$C = \bar{y}_{c^+} - \bar{y}_{c^-} = \frac{c + ac + bc + abc - a - b - ab - (1)}{4n}$$

L'interazione fra i due fattori  $AB$  è proprio la differenza nell'effetto di  $A$  ai due livelli di  $B$ , quando  $C$  è al livello basso, ed analogamente, quando  $C$  al livello alto:

$$AB(C \text{ basso}) = \frac{ab - b}{2n} - \frac{a - (1)}{2n} \qquad AB(C \text{ alto}) = \frac{abc - bc}{2n} - \frac{ac - c}{2n}$$

cosicché l'interazione  $AB$  è la media di queste due componenti, ovvero la differenza delle medie su due piani diagonali nel cubo:

$$AB = \frac{(1) + c + ab + abc - a - b - ac - bc}{4n}$$

Usando un approccio analogo, si trovano le stime degli effetti delle interazioni  $AC$  e  $BC$ :

$$AC = \frac{(1) + b + ac + abc - a - c - ab - bc}{4n}$$

$$BC = \frac{(1) + a + bc + abc - b - c - ab - ac}{4n}$$

Infine l'effetto dell'interazione  $ABC$  è la differenza delle medie delle interazioni  $AB$ , ai due livelli di  $C$ :

$$ABC = \frac{c + abc - ac - bc}{4n} - \frac{(1) + ab - a - b}{4n} = \frac{abc - ab - ac - bc + a + b + c - (1)}{4n}$$

I numeratori delle suddette equazioni sono detti contrasti, nelle otto combinazioni date dai livelli di fattori. e possono essere ricavati dalla tabella dei segni più e meno.

I segni per gli effetti principali (colonne  $A$ ,  $B$  e  $C$ ) si ottengono associando un segno più, al livello alto, ed un segno meno, al livello basso. Una volta ottenuti i segni per gli effetti principali, quelli per le altre colonne si determinano moltiplicando, riga per riga, le colonne degli effetti principali appropriati. La tabella ha alcune proprietà interessanti:

- eccetto che per la colonna identità  $I$ , ogni colonna ha egual numero di segni più e segni meno;
- la somma dei prodotti dei segni di ogni coppia di colonne è uguale a zero, perché le colonne sono ortogonali fra loro;
- il prodotto di ogni colonna per la colonna  $I$  dà la colonna stessa; perché la colonna  $I$  è l'identità;

- il prodotto di ogni colonna moltiplicata per se stessa è la colonna identità  $I$ .

		<i>Effetti fattoriali</i>								
		$I$	$A$	$B$	$C$	$AB$	$AC$	$BC$	$ABC$	
<i>Combinazioni di trattamenti</i>	1	(1)	+	-	-	-	+	+	+	-
	2	$a$	+	+	-	-	-	-	+	+
	3	$b$	+	-	+	-	-	+	-	+
	4	$c$	+	-	-	+	+	-	-	+
	5	$ab$	+	+	+	-	+	-	-	-
	6	$ac$	+	+	-	+	-	+	-	-
	7	$bc$	+	-	+	+	-	-	+	-
	8	$abc$	+	+	+	+	+	+	+	+

La stima di ogni effetto principale od interazione si determina moltiplicando la combinazione di livelli di fattori, elencati nella seconda colonna, per i segni della colonna corrispondente ad un certo effetto principale od una certa interazione, addizionando poi tutti gli addendi e dividendo il contrasto per la metà del numero totale di prove dell'esperimento:

$$effetto = \frac{contrasto}{n2^{k-1}}$$

cosicché la somma dei quadrati di ogni effetto risulta:

$$SS = \frac{(contrasto)^2}{n2^k}$$

### 9.5.3 Il piano $2^k$ senza replicazioni

In un esperimento fattoriale, all'aumentare del numero dei fattori, cresce anche il numero degli effetti stimati.

- Ad esempio, un esperimento  $2^4$  ha 4 effetti principali, 6 interazioni fra due fattori, 4 interazioni fra tre fattori ed 1 interazione fra quattro fattori.
- Invece un esperimento  $2^5$  ha 5 effetti semplici, 10 interazioni fra due fattori, 10 interazioni fra tre fattori, 5 interazioni fra quattro fattori ed 1 interazione fra cinque fattori.
- Inoltre un esperimento  $2^6$  ha 6 effetti semplici, 15 interazioni fra due fattori, 20 interazioni fra tre fattori, 15 interazioni fra quattro fattori, 6 interazioni fra cinque fattori ed 1 interazione fra sei fattori.

In molte applicazioni, si applica il principio di economia degli effetti, perché spesso un sistema è controllato dagli effetti principali e dalle interazioni di ordine basso, e le interazioni fra tre o più fattori sono trascurabili (cosicché quando si ha  $k \geq 4$  o  $5$ , è pratica comune eseguire il piano  $2^k$  senza interazioni o solo con le interazioni fra due fattori, conglobando le stime delle interazioni di ordine alto con la stima dell'errore.

#### 9.5.4 Aggiunta di punti centrali al piano $2^k$

Un potenziale problema nell'uso dei piani fattoriali, a due livelli, è l'ipotesi di non linearità negli effetti dei fattori. Ovviamente una linearità perfetta non è necessaria ed i piani  $2^k$  funzionano piuttosto bene anche quando l'ipotesi di linearità è verificata solo approssimativamente. Infatti quando un termine di interazione è aggiunto, ad un modello di soli effetti principali, è introdotta una curvatura nella superficie di risposta. Dato che il piano  $2^k$  supporta un modello con effetti principali ed interazioni, una certa forma di curvatura, nel piano, è inclusa comunque. Tuttavia in alcuni sistemi o processi, è necessario includere effetti del secondo ordine, al fine di ottenere un modello più adeguato. Considerato il caso  $k = 2$  fattori, un modello, includente effetti del secondo ordine, ha espressione:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

dove i coefficienti  $\beta_{11}$  e  $\beta_{22}$  misurano gli effetti quadratici puri. L'equazione costituisce un modello di superficie di risposta del secondo ordine. Questo modello non può essere adottato usando un piano di prove  $2^2$ , perché un modello quadratico richiede tutti i fattori, presenti nelle prove, con almeno tre livelli. D'altra parte, è importante, determinare, se i termini quadratici puri dell'equazione sono necessari.

Esiste un metodo, basato sull'aggiunta di un piano al piano fattoriale  $2^k$ , che può fornire qualche protezione contro gli effetti quadratici puri, nel senso che è possibile eseguire un test per determinare se questi effetti sono necessari. Inoltre se questo punto è replicato, si può ottenere una stima indipendente dell'errore atteso sperimentale. Questo metodo consiste nell'aggiungere punti centrali al piano  $2^k$ , costituiti da  $n_c$  repliche nel punto  $x_i = 0$  con  $i = 1, 2, \dots, k$ . Una ragione importante per aggiungere questi punti al centro è che non interferiscono con le usuali stime degli effetti del piano  $2^k$ .

Per illustrare questo approccio, si considera il piano  $2^2$  con un'osservazione in ciascuno dei punti fattoriali  $(-, -)$ ,  $(+, -)$ ,  $(-, +)$  e  $(+, +)$ , e con  $n_c$  repliche del punto centrale  $(0, 0)$ . Ovviamente si suppone che  $k$  fattori siano quantitativi, altrimenti il livello di mezzo o centrale non può esistere.

A riguardo, sia  $\bar{y}_f$  la media delle quattro prove, nei punti fattoriali, e  $\bar{y}_c$  la media delle  $n_c$  osservazioni nel punto centrale, se la differenza  $\bar{y}_f - \bar{y}_c$  è piccola, allora il punto centrale giace sul (oppure vicino al) piano, passante attraverso i piani fattoriali, e non si ha alcuna curvatura. Viceversa se  $\bar{y}_f - \bar{y}_c$  è grande, allora si è in presenza di curvatura. La somma di quadrati, con un grado di libertà per la curvatura quadratica pura e dove, in generale,  $n_f$  è il numero di punti nel piano fattoriale., è data da:

$$SS_{quadratica} = \frac{n_f n_c (\bar{y}_f - \bar{y}_c)^2}{n_f + n_c}$$

Pertanto questa quantità può essere confrontata con la media dei quadrati dell'errore, per eseguire un test sulla presenza o meno della curvatura.

In generale, quando si aggiungono punti al centro di un piano  $2^k$ , il modello diventa:

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{i=1}^{k-1} \sum_{j=i+1}^k \beta_{ij} x_i x_j + \sum_{j=1}^k \beta_{jj} x_j^2 + \varepsilon$$

dove i coefficienti  $\beta_{jj}$  sono gli effetti quadratici puri, cosicché il test per la curvatura verifica le ipotesi:

$$H_0: \sum_{j=1}^k \beta_{jj} = 0 \qquad H_1: \sum_{j=1}^k \beta_{jj} \neq 0$$

Inoltre se i punti del piano fattoriale non sono replicati, è possibile utilizzare gli  $n_c$  punti centrali, per costruire una stima dell'errore, con  $n_c - 1$  gradi di libertà.

### 9.5.5 Blocchi e confondimento nei piani $2^k$

E' spesso impossibile eseguire tutte le prove di un piano fattoriale  $2^k$ , in condizioni costanti e omogenee. Ad esempio, potrebbe non essere fattibile eseguire tutte le prove in un unico blocco o turno di lavoro, od usando solo il materiale prodotto nelle stesse condizioni. Quando si ha questo problema, il bloccaggio costituisce un'eccellente tecnica, per l'eliminazione delle variazioni indesiderate che possono essere causate dalle condizioni non omogenee. Se il piano contiene repliche e se i blocchi sono sufficientemente grandi, allora un metodo consiste nell'eseguire ogni replicazione in un unico blocco (quando un blocco è un insieme di condizioni omogenee). Ad esempio, si consideri un piano  $2^3$ , replicato due volte, supponendo necessaria circa un'ora, per completare una delle prove, ed eseguendo otto prove di ogni replicazione (una in un giorno e l'altra in un altro), si può eliminare ogni effetto temporale, cioè l'effetto del modo differente in cui il processo funziona nei due giorni. In questo modo, i due giorni diventano i due blocchi del piano sperimentale e la differenza media, fra le risposte nei due giorni, è pertanto l'effetto del blocco.

Invece a volte, è impossibile eseguire una replicazione completa di un piano fattoriale, sotto condizioni sperimentali omogenee. Si chiama confondimento una tecnica di pianificazione sperimentale, per eseguire esperimenti fattoriali in blocchi, quando la dimensione del blocco è inferiore alla numerosità di una replicazione completa del piano fattoriale. La tecnica fa sì che alcune interazioni siano indistinguibili o confuse con i blocchi.

Ad esempio, per quanto riguarda il confondimento, nel caso di piani fattoriali  $2^k$  in  $2^p$  blocchi, con  $p < k$ , dato un piano  $2^2$  (per ragioni di semplicità), si suppone che ognuna delle  $2^2 = 4$  prove richieda quattro ore, di analisi di laboratorio, che siano così necessari due giorni, per completare l'esperimento, e se i giorni sono considerati blocchi, che occorra assegnare due prove ad ogni giorno.

La Figura 9.5.2 mostra la rappresentazione geometrica del piano  $2^2$  e l'assegnazione delle prove ai blocchi, dove il blocco uno contiene le prove (1) ed  $ab$  ed il blocco due le prove  $a$  e  $b$ , cosicché i contrasti per stimare gli effetti di  $A$  e di  $B$  sono:

$$\text{contrasto}_A = ab + a - b - (1) \qquad \text{contrasto}_B = ab + b - a - (1)$$

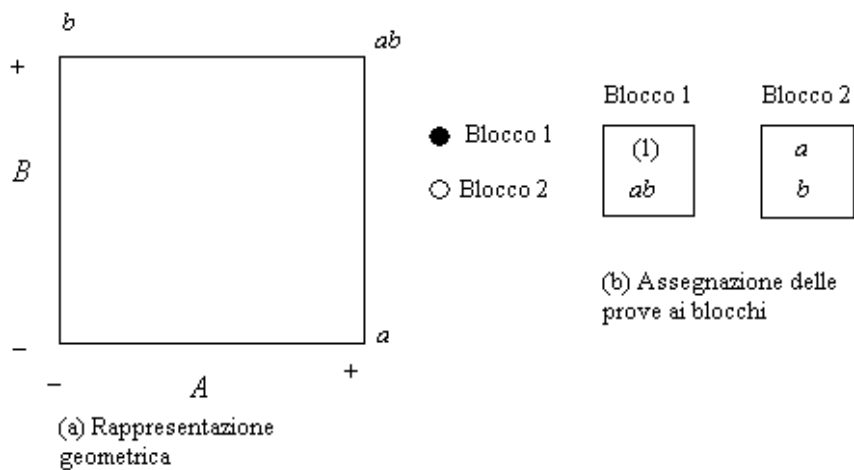


Fig. 9.5.2 – Un piano  $2^2$  in due blocchi

Questi contrasti non sono influenzati dal bloccaggio, dato che compaiono una prova con segno positivo ed una prova con segno negativo, da ciascun blocco, in ciascun contrasto ed contrasto per l'interazione  $AB$  è:

$$\text{contrasto}_{AB} = ab + (1) - a - b$$

In questo modo, dato che le prove con il segno positivo ( $ab$  ed  $(1)$ ) sono nel blocco uno, mentre quelle con il segno negativo ( $a$  e  $b$ ) sono nel blocco due, gli effetti del blocco e dell'interazione  $AB$  sono identici, cioè  $AB$  è confusa con i blocchi. Infatti nella tabella dei segni, per il piano  $2^2$ , le prove, con segno positivo, sono assegnate al blocco uno, mentre quelle, con segno negativo, sono assegnate al blocco due.

Questo schema è utile per confondere ogni piano  $2^k$  in due blocchi; ad esempio, dato un piano  $2^3$ , eseguito in due blocchi e supposto di voler confondere l'interazione fra i tre fattori  $ABC$  con i blocchi, dalla tabella dei segni, per il piano  $2^3$ , si assegnano le prove con segno meno, su  $ABC$ , al blocco uno e quelle con segno più, su  $ABC$ , al blocco due, come mostra la Figure 9.5.3.

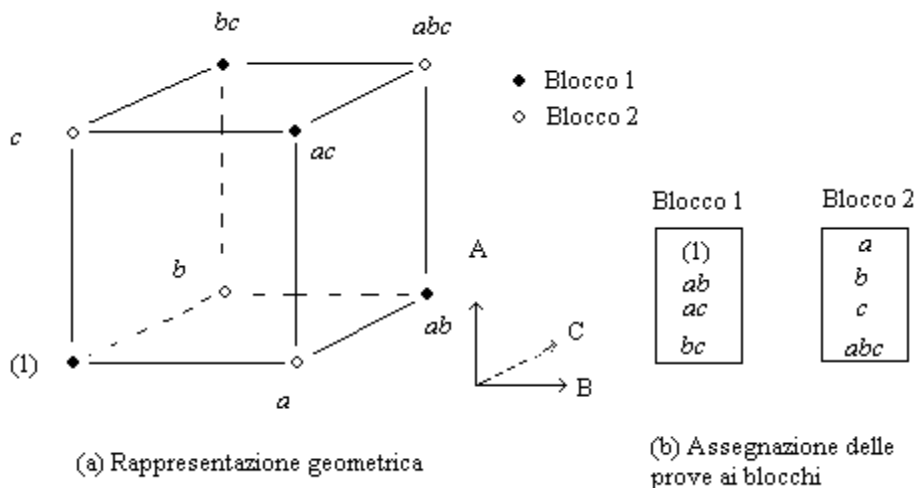


Fig. 9.5.3 – Un piano  $2^3$  in due blocchi con  $ABC$  confuso

## **10 Piani di campionamento per attributi lotto per lotto**

Il controllo o l'ispezione dei prodotti è uno degli aspetti dell'assicurazione della qualità. Quando il controllo o l'ispezione ha lo scopo di accettare o rifiutare il prodotto, in base alla corrispondenza agli standard richiesti, il genere di procedura impiegata è comunemente indicato come campionamento in accettazione. Il concetto soggiacente all'impiego ed alla progettazione dei piani di campionamento in accettazione è che essi non si sostituiscono al controllo del processo. Infatti un uso adeguato delle tecniche di controllo di processo, allo stadio iniziale del processo, comprendenti l'implementazione di opportuni controlli statistici, riduce drasticamente e talvolta elimina l'esigenza di altre ispezioni campionarie.

Questi piani di campionamento in accettazione, per attributi lotto per lotto, comprendono il progetto e la predisposizione di singoli piani di campionamento, l'utilizzazione della curva operativa caratteristica ed i concetti di ispezione e rettifica, di qualità media risultante e di numerosità media totale. Analoghi concetti sono introdotti per tipi di piani di campionamento in cui si considera più di un campione, per determinare l'accettabilità di un lotto (con un campionamento doppio o multiplo e sequenziale).

Infine altri sistemi di piani di campionamento standard sono progettati secondo due filosofie differenti e sono rispettivamente focalizzati sul livello di una qualità accettabile, riferita tanto a variabili, quanto ad attributi, ed orientati invece all'accertamento della percentuale di elementi difettosi del lotto, con la valutazione della qualità media del lotto in uscita.

### **10.1 Il problema del campionamento in accettazione**

il campionamento in accettazione si collega all'ispezione ed alla presa di decisione, in merito ai prodotti, ed è uno degli aspetti più antichi e degli elementi principali nel campo del controllo statistico della qualità. In anni recenti, è usuale lavorare per aumentare le prestazioni del prodotto, utilizzando lo SPC e la programmazione degli esperimenti, così da non essere dipendenti dal campionamento in accettazione, come strumento primo dell'assicurazione della qualità.

In una tipica applicazione del campionamento in accettazione, ricevuto un approvvigionamento di prodotti, questi sono spesso componenti o materiali grezzi, utilizzati nel processo di lavorazione. Dal lotto, è preso un campione e sono valutate alcune caratteristiche di qualità delle sue unità +. Sulla base delle informazioni ottenute da questo campione, si prende una decisione relativamente alla qualità del lotto. Di solito, questa decisione è accettare o rifiutare il lotto. I lotti accettati sono messi in produzione, mentre i lotti rifiutati sono rinviati, oppure sottoposti ad alcune altre azioni di accertamento.

Mentre si è soliti considerare il campionamento in accettazione come un'attività d'ispezione nel momento di ricevimento, esistono altri modi d'impiego dei metodi di campionamento. Ad esempio, sovente si campiona e si ispeziona il prodotto nelle diverse fasi di produzione. I lotti, così accettati, sono inviati ad ulteriori processi di lavorazione, mentre quelli rifiutati sono scartati o rilavorati.

A riguardo, tre aspetti del campionamento sono particolarmente importanti.

- Lo scopo del campionamento in accettazione è saggiare i lotti, non di stimare la qualità del lotto, e la maggior parte dei piani di campionamento in accettazione non è progettata per scopi di stima.
- I piani di campionamento in accettazione non forniscono alcuna forma diretta di controllo della qualità; in quanto, il campionamento in accettazione semplice accetta o rifiuta i lotti (ed anche qualora i lotti siano della medesima qualità, il campionamento ne accetta alcuni e ne rifiuta altri, pur non essendo i lotti

accettati migliori di quelli rifiutati: infatti il controllo di processo è utilizzato per controllare e migliorare sistematicamente la qualità, mentre non lo è il campionamento in accettazione);

- ❑ L'uso principale del campionamento in accettazione è valutare la qualità nel prodotto, ma come uno strumento di certificazione ed assicurazione che il risultato di un processo sia conforme ai requisiti richiesti.

In generale, esistono tre modi di intraprendere l'accertamento del lotto.

- ❑ L'accettazione senza ispezione, utile in situazioni in cui il processo produttivo è così buono che non sono quasi mai trovate unità difettose, oppure quando non si ha convenienza a cercare unità difettose.
- ❑ L'ispezione del cento per cento del lotto, ovvero l'ispezione di ogni unità del lotto, eliminando le unità difettose trovate, usata nei casi in cui il prodotto è estremamente critico ed accettare anche una sola unità difettosa determina un inaccettabile costo elevato per guasti, presenti nelle fasi successive (inoltre quando la capacità di processo è inadeguata a rispettare le specifiche richieste).
- ❑ il campionamento in accettazione utile quando:
  - ❑ la verifica è distruttiva;
  - ❑ il costo dell'ispezione al cento per cento è estremamente elevato;
  - ❑ l'ispezione al cento per cento non è tecnicamente attuabile, oppure richiede così tanto tempo che la programmazione ne risente seriamente;
  - ❑ si hanno molte unità da ispezionare ed il tasso di errore ispettivo è così alto per cui l'ispezione al cento per cento dà luogo all'accettazione di una percentuale di unità difettose più elevata di quanto accade, utilizzando un piano campionario;
  - ❑ si ha già un'eccellente immagine di qualità e si preferisce ridurre l'ispezione al cento per cento, ma la capacità è così bassa, da rendere l'assenza d'ispezione un'alternativa non soddisfacente;
  - ❑ esistono potenzialmente seri rischi di responsabilità e, benché il processo sia soddisfacente, si ritiene necessario un programma di continua osservazione del prodotto.

### 10.1.1 Vantaggi e svantaggi del campionamento

Il campionamento in accettazione, posto a confronto con l'ispezione al cento per cento, presenta alcuni vantaggi:

- ❑ è sovente meno costoso, perché si hanno meno ispezioni;
- ❑ si ha maneggiamento del prodotto ed i danneggiamenti sono ridotti;
- ❑ si può applicare nei casi di verifica distruttiva;
- ❑ un minor numero di persone è coinvolto nelle attività ispettive;
- ❑ spesso riduce la numerosità degli errori dell'ispezione;
- ❑ il rifiuto di interi lotti, invece del semplice rinvio dei pezzi difettosi, spesso induce una motivazione più forte a migliorare la qualità.

Tuttavia il campionamento in accettazione ha anche alcuni svantaggi:

- ❑ esiste il rischio di accettare lotti cattivi e rifiutare lotti buoni;
- ❑ in generale, si ottengono meno informazioni circa il prodotto od il processo di lavorazione;
- ❑ il campionamento in accettazione richiede, al contrario dell'ispezione al cento per cento, un'accurata pianificazione ed una documentazione completa della procedura di campionamento.

Il campionamento in accettazione è un territorio intermedio tra i due estremi dell'ispezione al cento per cento e della non ispezione, capace spesso di fornire una metodologia, per orientarsi tra i due estremi, ottenendo un grado sufficiente d'informazione sul controllo del processo. Infatti mentre non si ha un controllo diretto della qualità, nell'applicazione di un piano di campionamento in accettazione ad un lotto isolato, quando questo piano è applicato ad una serie di lotti, diventa un mezzo per ottenere una protezione sia per il committente che l'esecutore. Inoltre fornisce un insieme di informazioni sulla qualità, riguardanti il processo, e può produrre una retroazione, utile nel controllo del processo, tale da essere determinante qualora i controlli precedenti del processo non siano adeguati.

### 10.1.2 Tipi di piani di campionamento

Numerosi sono i modi per classificare i piani di campionamento in accettazione ed una delle classificazioni più invalsi è quella per attributi e per variabili. Le variabili sono le caratteristiche, misurate con una scala numerica, e gli attributi sono le caratteristiche, espresse sulla base della presenza o non presenza di difetti.

- ❑ Un piano di campionamento semplice è una procedura d'indagine di un lotto in cui è selezionato, in modo casuale, dal lotto un campione di  $n$  unità, cosicché la condizione del lotto è determinata sulla base dell'informazione contenuta solo in questo campione. Ad esempio, un piano di campionamento semplice per attributi consiste in una numerosità di campionamento  $n$  ed in un numero di accettazione minimo  $c$ , cosicché selezionate  $n$  unità, a caso, dal lotto, se si trova un numero di difetti nel campione minore od uguale a  $c$ , si accetta il lotto, mentre se questo numero è maggiore di  $c$ , il lotto è rifiutato.
- ❑ Un piano di campionamento doppio è più complesso. Infatti dopo aver selezionato un campione iniziale, è presa una decisione alternativa, basata sulle informazioni ricavate da questo campione, consistente in:
  - ❑ accettare il lotto,
  - ❑ rifiutare il lotto,
  - ❑ prelevare un secondo campione.

e, solo in questo ultimo caso, le informazioni, derivanti dal primo e dal secondo campione, sono unite, allo scopo di raggiungere una decisione sull'accettazione o sul rifiuto del lotto.

- ❑ Un piano di campionamento multiplo estende il concetto di piano di campionamento doppio, potendosi richiedere più di due campioni, per ottenere la decisione relativa alla situazione del lotto. Pertanto la numerosità del campione multiplo è spesso ridotta rispetto al campionamento semplice ed a quello doppio. Addirittura nel campionamento sequenziale, le unità sono prelevate una alla volta e, in seguito all'ispezione di ciascuna unità, è presa una decisione sull'accettazione od il rifiuto del lotto, ovvero sul prelevamento di un'altra unità.

### 10.1.3 La formazione del lotto

Il modo in cui il lotto per l'ispezione è costituito può influenzare l'efficacia del piano di campionamento in accettazione.

- ❑ I lotti devono essere omogenei le unità del lotto dovrebbero essere prodotte dalle stesse macchine, dagli stessi operatori e con i materiali grezzi uguali, approssimativamente nello stesso periodo di tempo. Quando i lotti non sono omogenei, come quando si mischiano produzioni di differenti linee, lo schema di campionamento in accettazione può non funzionare come di fatto dovrebbe. Lotti non omogenei rendono anche più difficile effettuare azioni correttive, per eliminare la fonte di prodotti difettosi.
- ❑ Lotti più grandi sono preferibili rispetto ai lotti più piccoli; infatti in generale, è più efficace ispezionare lotti grandi piuttosto che lotti piccoli.
- ❑ I lotti devono essere conformi al sistema di trattamento dei materiali utilizzato. In aggiunta, le unità dei lotti devono essere confezionate in modo tale da minimizzare i rischi di trasporto e di trattamento e da rendere relativamente facile la selezione delle unità per il campionamento.

### 10.1.4 Il campionamento casuale

Le unità selezionate dal lotto per l'ispezione devono essere scelte a caso e devono essere rappresentative di tutti gli elementi del lotto, ed il concetto casualità è estremamente importante per il campionamento in accettazione. A riguardo, per progettare un campionamento casuale, spesso la tecnica suggerisce di assegnare prima un numero a ciascuna unità del lotto e poi di individuare  $n$  numeri casuali, dove  $n$  va da uno al numero massimo di unità presenti nel lotto. Questa sequenza di numeri casuali determina quali unità del lotto servono a costituire il campione; infatti non si può prescindere dall'importanza del campionamento casuale e, se sono utilizzati metodi che non rispettano le condizioni di casualità per la scelta del campione, non valgono più le basi statistiche della procedura di campionamento in accettazione.

## 10.2 Piani di campionamento semplice per attributi

### 10.2.1 Definizione di un piano di campionamento semplice

Dato un lotto di dimensione  $N$ , sottoposto a ispezione, un piano di campionamento semplice è definito mediante la numerosità  $n$  del campione ed il numero di accettazione  $c$ . Dato poi il numero di elementi non conformi o difettosi osservati  $d$ :

- ❑ se questo numero di elementi non conformi o difettosi è minore o uguale a  $c$ , il lotto è accettato;
- ❑ se invece questo numero di elementi non conformi o difettosi osservati è maggiore di  $c$ , il lotto è rifiutato.

Dal momento che la caratteristica ispezionata è un attributo, ciascuna unità del campione può essere valutata come conforme o non conforme. In generale, uno o più tipi di attributo possono essere ispezionati, nel medesimo campione; e si definisce unità difettosa l'unità che non è conforme alle specifiche di uno o più attributi. Questa procedura è chiamata piano di campionamento semplice, poiché il lotto è valutato sulla base dell'informazione ricavata da un solo campione, di dimensione  $n$ .

### 10.3 Piani di campionamento doppio, multiplo e sequenziale

#### 10.3.1 Piani di campionamento doppio

Un piano di campionamento doppio è una procedura definita mediante quattro parametri in cui, in alcune situazioni, è richiesto un secondo campione, per valutare un lotto:

- ❑  $n_1$  dimensione del primo campione;
- ❑  $c_1$  numero di accettazione del primo campione;
- ❑  $n_2$  dimensione del secondo campione,
- ❑  $c_2$  numero di accettazione del secondo campione.

In questo modo, da un lotto è estratto un primo campione casuale, dove si riscontra un numero  $d_1$  di elementi non conformi o difettosi. Allora se  $d_1 \leq c_1$ , il lotto è accettato dopo il primo campionamento; invece se  $c_1 \leq d_1 \leq c_2$ , dallo stesso lotto si estrae un secondo campione casuale, riscontrando un numero  $d_2$  di elementi non conformi o difettosi, in questo secondo campione. Pertanto se  $d_1 + d_2 \leq c_2$ , il lotto è accettato; al contrario, se  $d_1 + d_2 > c_2$ , il lotto è rifiutato. La procedura di questo piano di campionamento è illustrata nella Figura 10.3.1,

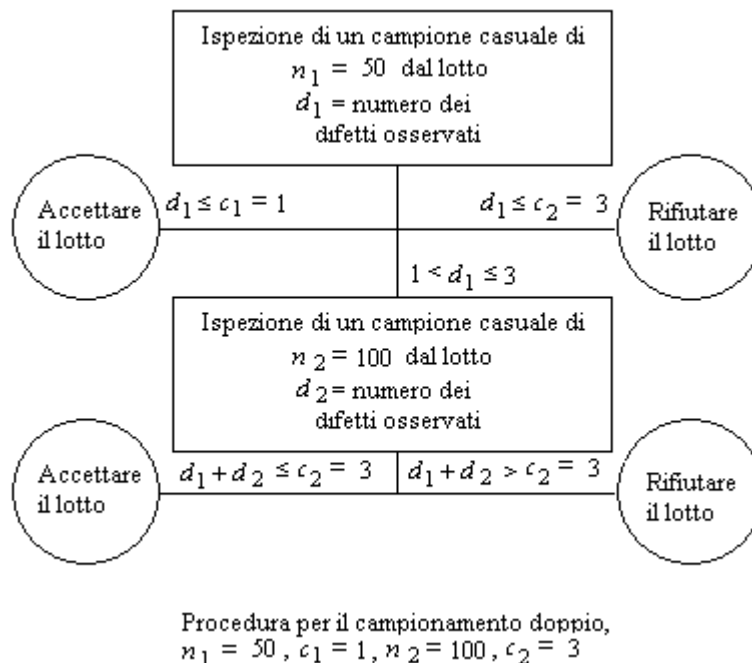


Fig. 10.3.1 - Piani di campionamento doppio

#### 10.3.2 Piani di campionamento multiplo

Un piano di campionamento multiplo è un'estensione del piano di campionamento doppio nel quale possono essere richiesti più di due campioni per decidere sul lotto. Questo piano opera nel modo seguente:

- ❑ se al completamento di ciascuna fase di campionamento il numero degli elementi non conformi o difettosi è minore od uguale al numero di accettazione, il lotto è accettato;
- ❑ se invece, nel corso di ogni fase, il numero di elementi difettosi supera il numero di rifiuto il lotto è rifiutato; altrimenti si estrae un altro campione.

Questa procedura multipla può continuare fino a che si estrae l'ultimo campione (avendo prefissato il numero corrispondente all'ultimo), dopodiché deve essere presa una decisione in merito all'intero lotto. Di solito, il primo campione è ispezionato al cento per cento, mentre i successivi campionamenti sono spesso soggetti a procedura abbreviativa.

Il vantaggio principale dei piani di campionamento multiplo consiste nel fatto che i campioni, richiesti ad ogni stadio, sono spesso più piccoli di quelli richiesti dal campionamento semplice o doppio, cosicché questa procedura può comportare vantaggi, benché operativamente molto più complessa.

### **10.3.3 Piani di campionamento sequenziale**

Il campionamento sequenziale è un'estensione del campionamento doppio e multiplo. Nel campionamento sequenziale infatti si prende una successione di campioni dal lotto e si fa sì che il numero di campioni sia interamente determinato dai risultati del processo di campionamento. Il campionamento sequenziale può in via teorica continuare indefinitamente, finché il lotto non sia ispezionato al cento per cento. In pratica, i piani di campionamento sequenziale sono spesso interrotti, quando il numero delle ispezioni è uguale a tre volte il numero delle ispezioni che si sarebbero effettuate utilizzando un corrispondente piano di campionamento semplice. Se la dimensione del campione ispezionato, a ogni fase, è uguale a uno, la procedura è detta campionamento sequenziale unitario.

## **10.4 Procedure di campionamento per attributi mediante piani standard**

### **10.4.1 Descrizione della normativa**

La procedura di campionamento mediante piani standard, per l'ispezione per attributi, è il sistema di piani di accettazione più ampiamente usato oggi nel mondo. La versione originale delle tavole di campionamento standard risale al 1950; da allora, si sono fatte quattro revisioni e l'ultima versione è del 1989.

Uno schema di campionamento è una strategia complessiva, descritta in una normativa e composta da una collezione di schemi di campionamento, per specificare il modo in cui i piani di campionamento devono essere utilizzati. Le tavole di campionamento standard forniscono tre tipi di campionamento: semplice, doppio e multiplo. Per ciascun tipo di piano di campionamento, è fornito il materiale:

- ❑ per l'ispezione normale, impiegata all'inizio dell'attività ispettiva;
- ❑ per quella rinforzata, intrapresa quando la casistica recente della qualità presenta elementi negativi (in questo caso, i requisiti ispettivi più ristrettivi rispetto a quelli dell'ispezione normale);
- ❑ per quella ridotta, in generale, usata nel caso di un'ispezione inferiore a quella per l'ispezione normale.

Il primo punto fondamentale è il livello di qualità accettabile (AQL) e la tavola di campionamento standard è indicizzata con riferimento ad una serie di AQL la cui ampiezza va da 0.1% a 10%.

I livelli AQL sono collocati in progressione e ciascun livello è approssimativamente pari ad 1.585 volte il precedente. Ai livelli minori di AQL, lo stesso piano di campionamento può essere usato per controllare la frazione di elementi non conformi o difettosi, come pure il numero di difetti per unità.

La dimensione del campione usata è data dalla dimensione del lotto e dalla scelta del livello d'ispezione, essendo tre i livelli generali d'ispezione forniti:

- il secondo livello è indicato come normale;
- il primo livello richiede circa la metà dell'ammontare delle ispezioni del secondo livello e può essere usato, quando si necessita di una minore discriminazione;
- il terzo livello richiede circa il doppio delle ispezioni del secondo livello e deve essere utilizzato, quando si esige maggiore discriminazione.

Le procedure di passaggio tra le ispezioni normale, rinforzata e ridotta sono così descritte:

- da normale a rinforzata;
- da rinforzata a normale;
- da normale a ridotta;
- da ridotta a normale;
- discontinuità dell'ispezione.

#### **10.4.2 Descrizione della procedura**

La procedura di campionamento mediante piani standard è composta dai seguenti passi:

- scegliere il livello di qualità accettabile (AQL);
- scegliere il livello d'ispezione;
- determinare la dimensione del lotto;
- trovare la lettera di codice, appropriata per la dimensione del campione, nell'apposita tabella;
- determinare il tipo appropriato di piano di campionamento standard (semplice, doppio o multiplo),
- utilizzare la tabella appropriata, per individuare il piano da impiegare;
- determinare il corrispondente piano normale o ridotto, da usare per l'ispezione.

### **11. Altre tecniche di campionamento in accettazione**

Altre tecniche utili di campionamento in accettazione, comprendono i piani di campionamento per variabili, impiegati in alternativa ai piani per attributi, quando sono disponibili i dati di misurazione.

#### **11.1 Campionamento in accettazione per variabili**

##### **11.1.1 Vantaggi e svantaggi del campionamento per variabili**

Il principale vantaggio dei piani di campionamento per variabili consiste nel fatto che la stessa curva operativa caratteristica (OC) può essere ottenuta con una dimensione del campione ridotta, rispetto a quella richiesta da un piano di campionamento per attributi (cosicché un piano di campionamento per variabili, con

la stessa protezione statistica di un piano di campionamento per attributi, richiede un campionamento minore).

Un secondo vantaggio è che i dati di misurazione forniscono spesso una maggiore informazione sul processo o sul lotto, rispetto ai dati per attributi. Infatti in generale, le misure numeriche delle caratteristiche di qualità sono più utili della semplice classificazione degli elementi difettosi e non difettosi.

Infine quando i livelli di qualità accettabili sono piuttosto bassi, la dimensione del campione, richiesta dal campionamento per attributi, è molto grande e, in queste situazioni, può essere opportuno rivolgersi alla misurazione per variabili. Di conseguenza, poiché è molto importante avere numeri di elementi non conformi o difettosi in termini di parti per milione, il campionamento per variabili diventa vantaggioso.

### 11.1.2 Tipi di piani di campionamento disponibili

I due tipi di procedure di campionamento per variabili sono:

- i piani per controllare la frazione di elementi non conformi o difettosi del processo o del lotto;
- i piani per controllare i parametri (in generale, la media) del processo o del lotto.

A riguardo, dato un piano di campionamento per variabili per il controllo della frazione di elementi non conformi o difettosi del processo o del lotto, poiché la caratteristica di qualità è una variabile, si hanno un limite di specificazione inferiore (LSL) od un limite di specificazione superiore (USL), oppure entrambi, limiti che definiscono i valori accettabili di questo parametro.

Allora data una caratteristica di qualità  $x$ , normalmente distribuita, e supposto che, su questa grandezza, esista un limite di specificazione inferiore, il simbolo  $p$  indica la frazione di elementi non conformi o difettosi (dove la frazione di elementi non conformi o difettosi è funzione della media  $\mu$  e della deviazione standard  $\sigma$  del processo o del lotto).

Supposta nota la deviazione standard, sotto questa condizione, si vuole eseguire un campionamento del processo o del lotto, per determinare quale valore debba assumere la media affinché la frazione di elementi non conformi o difettosi  $p$  sia accettabile.

- Metodo  $k$

Dato un campione casuale di  $n$  elementi, estratto da un lotto, si calcoli la statistica:

$$Z_{LSL} = \frac{\bar{x} - LSL}{\sigma}$$

dove  $Z_{LSL}$  esprime la distanza tra la media del campione ed il limite inferiore di specificazione, in unità di deviazione standard. Quanto più elevato è il valore di  $Z_{LSL}$ , tanto più la media del campione dista dal limite inferiore di specificazione e tanto minore è la frazione  $p$  di elementi non conformi o difettosi.

Pertanto se è stabilito un valore critico di  $p$  da non superare, con una probabilità stabilita, si può trasformare il valore di  $p$  nella distanza critica  $k$ , per  $Z_{LSL}$ , cosicché:

- ❑ se  $Z_{LSL} \geq k$ , si deve accettare il lotto, poiché i dati del campione implicano che la media del lotto è sufficientemente al di sopra di LSL, assicurando così che la frazione di elementi non conformi o difettosi presenti nel lotto è soddisfacente.
- ❑ se invece  $Z_{LSL} < k$ , la media è troppo vicina al LSL ed il lotto deve essere rifiutato.

#### ❑ Metodo $M$

Dato un campione casuale di  $n$  elementi, estratto da un lotto, si calcoli ancora la statistica:

$$Z_{LSL} = \frac{\bar{x} - LSL}{\sigma}$$

utilizzando  $Z_{LSL}$ , per stimare la frazione di elementi non conformi o difettosi, presenti nel lotto, come area sottostante la curva normale standardizzata (dove l'impiego di  $Q_{LSL} = Z_{LSL} \sqrt{n/(n-1)}$  come statistica, con distribuzione normale standardizzata, è migliore, poiché fornisce una stima corretta di  $p$ ).

Dopodiché sia  $\hat{p}$  la stima di  $p$ , così ottenuta:

- ❑ se la stima  $\hat{p}$  è inferiore od uguale al valore massimo specificato  $M$ , il lotto è accettato;
- ❑ se invece la stima  $\hat{p}$  è superiore al valore massimo specificato  $M$ , il lotto non è accettato.

### 11.2 Descrizione della normativa

Anche è il piano campionario di accettazione, lotto per lotto, per variabili è descritto da una normativa, introdotta nel 1957, ed il suo punto centrale è ancora il livello di qualità accettabile (AQL), variabile tuttavia tra 0.04% e 15%. In questo caso, sono previsti sono cinque livelli d'ispezione:

- ❑ il quarto livello d'ispezione è definito come normale;
- ❑ il quinto livello d'ispezione ha una curva delle caratteristiche operative (OC) calcolata per un numero di punti maggiore rispetto a quello del quarto livello;
- ❑ i livelli d'ispezione inferiori, sono invece adottati, quando è necessario ridurre il campionamento e se si possono o si devono tollerare rischi maggiori.

Come nella normativa della procedura per attributi, sono usate lettere di codice, per definire la dimensione campionaria, ma la stessa lettera non implica stessa dimensione nelle due normative, ed anche le classi di dimensione del processo o del lotto sono diverse nelle due normative.

Le dimensioni campionarie sono in funzione della dimensione del lotto e del livello d'ispezione, anche in questo caso: normale, rinforzata e ridotta.

In ogni caso, sia i piani di campionamento che le procedure, nelle tavole di campionamento della normativa, si basano sul presupposto che il campionamento sia casuale e che la caratteristica di qualità d'interesse sia distribuita normalmente.

## APPENDICE B – TEORIA DELLA ROBUSTEZZA

La non – linearità della norma si manifesta nella funzione obiettivo che segna, così, un distacco definitivo dai minimi quadrati. D'altra parte, poiché, come ben noto, i sistemi lineari si possono risolvere con appropriati algoritmi numerici, mentre i sistemi non – lineari presentano, in generale, evidenti difficoltà, si sceglie o, comunque, si preferisce risolvere questi in modo iterativo, operando, di volta in volta, in ambito lineare.

Allora funzioni obiettivo, diverse da quelle dei minimi quadrati, sono ricondotte a queste, tramite la definizione di opportuni pesi ausiliari, per operare, iterativamente, in ambito lineare, con la procedura cosiddetta dei minimi quadrati ripesati. Si ricorda, per completezza, che è noto, in letteratura, il metodo del semplice, capace di risolvere sistemi con funzione obiettivo della minima somma dei moduli (come nella programmazione lineare, programmazione lineare vincolata, ecc.); tuttavia poiché tale metodo, proprio della ricerca operativa, è certamente più complesso e male si presta, comunque, alla soluzione di grandi sistemi, esso è considerato estraneo agli scopi del presente lavoro.<sup>4</sup>

La formalizzazione matematica della trasformazione di una funzione obiettivo qualsiasi in quella dei minimi quadrati coincide con la definizione stessa dei pesi ausiliari, capaci di operare questa trasformazione:

$$\phi = \|p\hat{v}\|_k = \|qp\hat{v}\| = \min \quad ; \quad q = \frac{1}{(|v|^{2-k} + \varepsilon)}$$

essendo  $k$  l'indice della norma,  $\hat{v}$  gli scarti–residui,  $p$  i pesi delle osservazioni (o pseudo–osservazioni),  $q$  i pesi ausiliari e  $\varepsilon$  un numero positivo piccolissimo, necessario quando, insieme, lo scarto–residuo è nullo e l'esponente (a denominatore) positivo.

La procedura, nota come minimi quadrati ripesati, collega il problema della non – linearità della norma a quello della non – linearità del modello stocastico, riportando la sua soluzione alla procedura, nota come riproduzione dei pesi, già descritta in precedenza.

Una prima soluzione ai minimi quadrati di tutte le osservazioni (o meglio di una parte di esse, la cui bontà sia da ritenere nota o conosciuta per altra via) avvia la procedura. Gli scarti–residui di tutte le equazioni concorrono a costruire i pesi ausiliari, in base ai quali effettuare una nuova iterazione.

La procedura continua, iterativamente, fino ad ottenere la convergenza del metodo con la stabilizzazione dei pesi ausiliari. Infine resta da sottolineare come questa procedura sia adatta, in particolare, all'uso di stimatori robusti, come noto, ben più capaci dei minimi quadrati di identificare ed eliminare dati anomali, la cui presenza ha, invece, effetti distorcenti, quando si opera con i minimi quadrati (infatti, come già detto, la robustezza è una proprietà antitetica al metodo dei minimi quadrati).

Le Figure B.1 e B.2 mostrano le funzioni obiettivo, le funzioni d'influenza (si veda, a riguardo, la prima appendice del presente lavoro) ed i pesi ausiliari, così come sono stati definiti e proposti, rispettivamente, da Huber e Hampel.

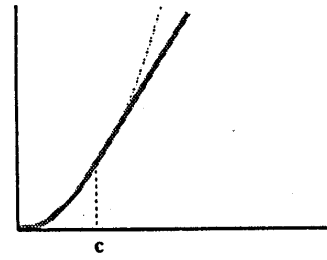
---

<sup>4</sup> Si chiama metodo del semplice un algoritmo capace di ricercare l'estremo (massimo e minimo) di una funzione lineare di più variabili, sottoposte a vincoli di disuguaglianza, anch'essi di tipo lineare. Da un punto di vista geometrico, l'insieme delle soluzioni possibili è un poliedro limitato dagli iperpiani (rette nel caso bidimensionale, piani nel caso tridimensionale), rappresentazione geometrica dei vincoli. Il metodo permette di passare da un vertice ad un altro del poliedro, alla ricerca di quello (o quelli), dove la funzione raggiunge il suo estremo. Pertanto la soluzione corrisponde sempre ad (almeno) uno dei vertici del poliedro.

## Stimatore di Huber

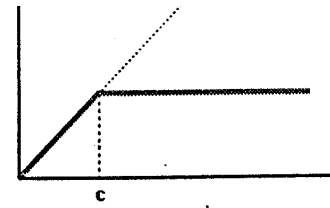
funzione obiettivo:

$$\phi(x) = \begin{cases} \frac{1}{2}x^2 & \|x\| < c \\ c\|x\| - \frac{1}{2}c^2 & \|x\| \geq c \end{cases}$$



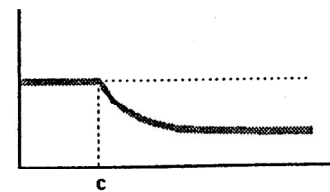
funzione d'influenza:

$$\psi(x) = \begin{cases} x & \|x\| < c \\ c \operatorname{sign}(x) & \|x\| \geq c \end{cases}$$



funzione dei pesi ausiliari:

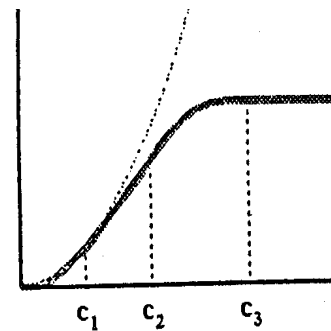
$$q(x) = \begin{cases} 1 & \|x\| < c \\ \frac{c}{\|x\|} & \|x\| \geq c \end{cases}$$



## Stimatore di Hampel

funzione obiettivo:

$$\phi(x) = \begin{cases} \frac{1}{2}x^2 & \|x\| < c_1 \\ c_1\|x\| - \frac{1}{2}c_1^2 & c_1 \leq \|x\| \leq c_2 \\ c_1c_2 - \frac{1}{2}c_1^2 + \frac{1}{2}c_1(c_3 - c_2) \left[ 1 - \left( \frac{c_3 - \|x\|}{c_3 - c_2} \right)^2 \right] & c_2 \leq \|x\| \leq c_3 \\ c_1c_2 - \frac{1}{2}c_1^2 + \frac{1}{2}c_1(c_3 - c_2) & \|x\| \geq c_3 \end{cases}$$



funzione d'influenza:

$$\psi(x) = \begin{cases} x & \|x\| < c_1 \\ c_1 \operatorname{sign}(x) & c_1 \leq \|x\| \leq c_2 \\ \frac{c_1}{c_3 - c_2} (c_3 - \|x\|) \operatorname{sign}(x) & c_2 \leq \|x\| \leq c_3 \\ 0 & \|x\| \geq c_3 \end{cases}$$

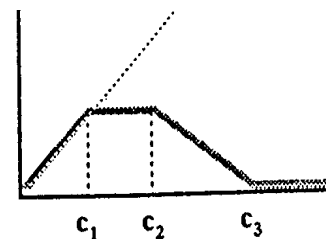


Fig. B.1

funzione dei pesi ausiliari:

$$q(x) = \begin{cases} 1 & \|x\| < c_1 \\ \frac{c_1}{\|x\|} & c_1 \leq \|x\| \leq c_2 \\ \frac{c_1}{(c_3 - c_2)} \left( \frac{c_3}{\|x\|} - 1 \right) & c_2 \leq \|x\| \leq c_3 \\ 0 & \|x\| \geq c_3 \end{cases}$$

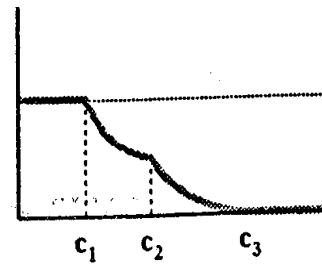


Fig. B.2

Lo stimatore di Hampel ha il pregio di eliminare automaticamente tutte quelle osservazioni che si allontanano dal grosso dei dati. Al contrario, lo stimatore di Huber rimane più vicino al comportamento dei minimi quadrati e, pur difettando un po' in efficacia della robustezza, conserva meglio le proprietà ottimali dei minimi quadrati. Si noti tuttavia come entrambi gli stimatori, essendo non – lineari, richiedano una prima soluzione ai minimi quadrati di tutte le osservazioni o di una parte di esse (la cui bontà sia nota o conosciuta per altra via) e come entrambe le strade presentino evidenti difficoltà. Infatti data la non robustezza dei minimi quadrati, tutte le osservazioni possono alterare molto i risultati attesi, mentre conoscere per altra via il grosso dei dati è spesso impresa assai difficile.

Un'alternativa vantaggiosa, proposta di recente, è il metodo della **minima mediana dei quadrati degli scarti – residui** (o identicamente del **minimo mav dei moduli degli scarti – residui**). Esso richiede l'esecuzione, in successione, di tanti sistemi quanti richiesti dal binomio di Newton <sup>5</sup>, aventi ciascuno un numero di equazioni pari al numero di incognite (cioè di parametri), scegliendo come soluzione quello che fornisce il minimo cercato. Poiché solitamente il numero di osservazioni eccede notevolmente il numero di parametri, il numero di sistemi da risolvere diventa rapidamente grande al crescere della ridondanza. Allora una strategia praticabile limita l'applicazione del metodo a piccoli sotto – sistemi.

Altre strategie praticamente computabili sono possibili adottando metodi campionari, come il **RANSAC** (cioè **Random Sampling**, ovvero **campionamento causale**), dove invece di calcolare tutte le combinazioni possibili, ricercando la minima mediana dei quadrati degli scarti – residui, si prova a risolvere un certo numero di sistemi <sup>6</sup>, ciascuno avente un numero di equazioni pari al numero delle incognite. Per procedere così, è necessario sorteggiare ogni volta, dall'insieme delle osservazioni, tante osservazioni quanti sono i parametri (incogniti) del sistema da risolvere. Dopodiché si costruisce l'istogramma delle soluzioni ottenute: la soluzione modale (ovvero quella con la maggior frequenza assoluta o presenza) è accettata, come stima probabile della suddetta minima mediana <sup>7</sup> (a riguardo, una convergenza del metodo è ragionevolmente attesa, pur in assenza di qualsiasi garanzia di convergenza asintotica).

<sup>5</sup> Binomio di Newton:  $\binom{m}{n} = \frac{m!}{(m-n)!n!} \rightarrow m=15 \quad n=5 \quad \binom{m}{n} \cong 3.000$ ;  $m=20 \quad n=7 \quad \binom{m}{n} \cong 80.000$ ;  $m=30 \quad n=10 \quad \binom{m}{n} \cong 30.000.000$ .

Si noti come detto binomio cresca maggiormente, se  $n$  è circa la metà di  $m$  e come, d'altra parte, utili considerazioni sul bilanciamento ottimale tra il numero di parametri o incognite ed il numero di osservazioni o equazioni consigli un terzo del primo rispetto al secondo.

<sup>6</sup> Il numero di tali sistemi è  $k = \frac{\ln \alpha}{\ln(1 - (1 - \epsilon)^n)}$ , essendo  $\alpha$  il livello di significatività del test ed  $\epsilon$  la percentuale massima di dati anomali.

<sup>7</sup> Evidenti controindicazioni al metodo, sono date dalla presenza di bi-modalità e multi – modalità. Al contrario, questo metodo è ulteriormente accelerato, se non – tutte le osservazioni / equazioni possono essere casualmente selezionate, per calcolare la soluzione del sistema da risolvere, perché non – tutte le incognite, corrispondenti ai parametri (del problema dato), sono presenti in tutte le equazioni. Di conseguenza, le scelte possibili si restringono, aumentando la computabilità, anche se difficile diventa, a volte, la scelta delle equazioni, specie nel caso di strutture reticolari.

I tre concetti fondamentali dell'approccio infinitesimale alla robustezza sono la robustezza qualitativa (collegabile al concetto intuitivo di robustezza, come stabilità o continuità dello stimatore, per piccoli cambiamenti nella funzione di distribuzione di probabilità sottostante), la funzione di influenza ed il punto di rottura. Il primo corrisponde alla continuità uniforme di una funzione, il secondo alla sua derivata, il terzo alla distanza, rispetto alla singolarità più vicina, del punto in cui si calcola la derivata.

- La **funzione di influenza**  $IF(X, T, F)$  consente di descrivere alcune proprietà, di notevole importanza, dello stimatore  $T$ . In particolare, il parametro caratterizzante l'estremo superiore della  $IF$ :

$$\gamma = \sup_x |IF(x, T, F)|$$

misura la **gross-error sensitivity** dello stimatore, ovvero, la massima deviazione causata da una contaminazione infinitesimale di  $F$ . Si noti che questa contaminazione può essere rappresentata, sia da un piccolo cambiamento in qualcuna o, addirittura, tutte le osservazioni (errori di arrotondamento, raggruppamento in classi dei dati), sia da forti variazioni in qualcuna soltanto (gross errors). Se  $\gamma$  è un valore finito, ovvero  $IF$  è limitata, lo stimatore è robusto (tanto più robusto, quanto minore è  $\gamma$ ).

Una seconda caratteristica è la presenza o meno di un **punto di reiezione** (in inglese, **rejection point**)  $x_0$ , ovvero un valore limite all'entità della contaminazione, per cui è nulla l'influenza di errori superiori ad un certo valore. In particolare, gli stimatori con funzione di influenza che va a zero, con continuità, (stimatori discendenti, di tipo  $S$ ) sono quelli con un meccanismo di forte sottopesatura, a posteriori, delle osservazioni contenenti gli errori più grandi.

- La funzione di influenza rappresenta un utile strumento, per comprendere il comportamento locale dello stimatore: infatti descrive la stabilità del valore asintotico di  $T$  per una variazione (contaminazione) infinitesimale di  $F$ . Tuttavia è altrettanto importante, avere un'idea del comportamento globale dello stimatore, rispetto a violazioni più consistenti del modello. Il **punto di rottura** (in inglese: **breakdown point**) indica, proprio, fino a quale distanza dal modello lo stimatore fornisce, ancora, informazioni attendibili. Si tratta, in sostanza, di capire quale frazione di dati anomali, non appartenenti al modello, può essere tollerata nel campione, da parte dello stimatore, prima di risultare non informativo. La nozione è, ancora, legata alla misura di distanza fra due funzioni di distribuzione di probabilità  $d^*(F, G)$  ed al fatto che, finché le funzioni di distribuzione di probabilità delle variabili casuali campionarie sono prossime fra loro:  $d^*(L_F(T_n), L_G(T_n)) < \varepsilon$ , la statistica  $T_n$  si mantiene, in probabilità, vicino al parametro (incognito). Dal punto di vista degli "outliers", si può definire, esplicitamente, un punto di rottura per il "gross error model":

$$F(x - \theta) = (1 - \varepsilon) G(x - \theta) + \varepsilon H(x - \theta)$$

per cui  $\varepsilon$  risulta, proprio, la frazione massima di "outliers" (cioè di dati anomali e/o di errori grossolani, veri e propri) ammessa nel campione.

## APPENDICE C – ESEMPI DI ANALISI SPAZIALE

L'analisi spaziale è uno strumento basilare per il trattamento delle osservazioni e l'analisi dei dati, in tutte le discipline del rilevamento, dalla geodesia alla cartografia con le loro scienze e tecniche affini, come pure oggi tanto in Geomatica, quanto in Geomatica Applicata, e si può estendere all'analisi spaziale di basi di dati dinamiche, con l'analisi spazio-temporale delle stesse.

### VARIABILE STATISTICA SEMPLICE<sup>8</sup>

#### DATI IN INGRESSO

VALORI OSSERVATI (NUMERO DI OSSERVAZIONI = 16)

-0.42 1.13 0.09 -2.01 0.77 0.00 -0.44 0.48 1.91 -0.18 -1.19 -0.24 0.56 0.98 -1.31 0.13

#### ELABORAZIONI E RISULTATI

##### STATISTICHE OTTIMALI

MEDIA	SCARTO QUADRATICO MEDIO (SQM o sqm)	ASIMMETRIA	CURTOSI
0.02	0.96	-0.20	2.81

##### STATISTICHE ROBUSTE

MEDIANA	MEAN ABSOLUTE VALUE (MAV)	MEDIAN ABSOLUTE VALUE (mav)
0.09	0.74	0.53

### VARIABILE STATISTICA DOPPIA

#### DATI IN INGRESSO

TABELLA A DOPPIA ENTRATA DEI VALORI OSSERVATI  
(NUMERO DI OSSERVAZIONI = 31)

X / Y	1	2	3	4	Marginale(Y)	Media(Y X)	Sqm(Y X)
1			9		9	3.0	0.0
2	3		2		5	1.8	1.0
3			4	4	8	3.5	0.5
4	1	1		7	9	3.4	1.1
Marginale(X)	4	1	15	11	31		
Media(X Y)	2.5	4.0	1.7	3.6			
Sqm(X Y)	0.9	0.0	0.9	0.5			

<sup>8</sup> La statistica descrittiva introduce allo studio della statistica ed è preliminare a qualsiasi altro studio più complesso ed avanzato, perché qualsiasi altra analisi dei dati, così come qualsiasi algoritmo e/o procedura della statistica computazionale, non può prescindere dalla conoscenza delle statistiche di base, proprie della statistica descrittiva. In particolare, accogliendo le tradizionali ipotesi di normalità dei dati, provenienti da campioni ripuliti, lo studio delle variabili statistiche e casuali (ovvero dei dati provenienti dagli esperimenti realmente eseguiti e dei modelli astratti usati per l'interpretazione degli stessi) semplici, cioè ad una dimensione, e doppie, cioè a due dimensioni, sono basilari. D'altra parte, poiché l'ipotesi di poter lavorare con campioni ripuliti deve necessariamente essere verificata ed approvata, data la inevitabile presenza di dati anomali (da intendersi, per lo più, come errori grossolani), statistiche monodimensionali e bidimensionali della robustezza devono affiancare le suddette e più efficienti statistiche della normalità. Maggiori informazioni ed i dettagli del calcolo sono presenti nel sito del Dipartimento di Ingegneria Civile e Ambientale (DICA) del Politecnico di Milano: <http://intranet.dica.polimi.it/people/mussio-luigi/> alla voce: [http://intranet.dica.polimi.it/uploads/media/2\\_validazione\\_dati.pdf](http://intranet.dica.polimi.it/uploads/media/2_validazione_dati.pdf).

## ELABORAZIONI E RISULTATI

$$\begin{array}{ll} M(X) = 2.55 & M(Y) = 3.06 \\ SQM(X) = 1.19 & SQM(Y) = 0.95 \end{array}$$

### ANALISI DELLA CONNESSIONE

INDICI DI BONFERRONI:

$$B(X) = 0.56 \quad B(Y) = 0.66 \quad B(-1) = 0.61 \quad B(0) = 0.61$$

### ANALISI DELLA REGRESSIONE

INDICI DI PEARSON:

$$ETA^2(X|Y) = 0.61 \quad ETA^2(Y|X) = 0.39 \quad ETA^2 = 0.53$$

### ANALISI DELLA CORRELAZIONE

RETTE DI REGRESSIONE:

$$\begin{array}{lll} A(Y(X)) = 0.25 & A(X(Y)) = 0.39 & \\ B(Y(X)) = 2.43 & B(X(Y)) = 1.35 & R(X,Y) = 0.31^9 \end{array}$$

Il breve sommario con l'elaborazione di due semplici esempi di statistica descrittiva, relativi a due variabili, rispettivamente semplice e doppia, fornisce anche i dati di ingresso per alcuni esempi di analisi spaziale, con metodologie dell'analisi multivariata, quali la cluster analysis, l'analisi di varianza (ANOVA) e la regressione multipla (nella modalità di una regressione temporale cubica e di una piana lineare).

Infatti queste metodologie di analisi multivariata sono impiegate, per lo più, prendendo in considerazione grandi basi di dati, perché grandi sono, quasi sempre, i problemi da affrontare nella gestione e nel controllo dell'ambiente e del territorio. Tuttavia l'estrema semplicità degli esempi proposti serve a mostrare l'effettiva praticabilità di tutte queste metodologie.

### CLUSTER ANALYSIS

X / Y	1	2	3	4
1			9	
2	3		2	
3			4	4
4	1	1		7

<sup>9</sup> Un commento sui pregi ed i limiti dello studio della dipendenza (muovendosi dalla connessione alla correlazione, attraverso la regressione o dipendenza funzionale) mette in evidenza come la variabile doppia dia informazioni maggiori della variabile semplice. Infatti se la variabile semplice dà informazioni sul centro e la dispersione di una certa variabile statistica, nonché indicazioni sulla forma della variabile casuale che può fornire un modello probabilistico di comportamento, la variabile doppia tratta essenzialmente della dipendenza. Questo analisi è di gran lunga più complessa e delicata, perché ha il compito di mettere in evidenza l'esistenza o meno di un legame statistico tra due variabili in esame e, nel caso della sua esistenza, sulla sua natura e la sua forza.

E' altrettanto evidente come queste informazioni siano insieme, quasi sempre, di grande interesse e, altrettanto spesso, assai incerte. Infatti studiare la dipendenza è di grande interesse, perché una variabile potrebbe essere più facilmente acquisibile (talvolta è il solo dato acquisibile) di un'altra maggiormente importante, oppure tracciante o precursore di un'altra che si vuole tenere sotto controllo preventivo (e così via con esempi consimili), mentre la criticità sta nel rischio di dare informazioni già con campioni non troppo grandi, data la grande incertezza delle stime di dipendenza per piccoli campioni (qualsiasi esse siano), come mostrano, ad esempio, le curve di potenza del coefficiente di correlazione.

L'analisi della disposizione spaziale delle informazioni comprese nella tabella a doppia entrata (che ha già condotto allo studio della dipendenza, in quanto variabile statistica doppia) permette di individuare tre diversi raggruppamenti di dati (in inglese, detti: cluster, cioè grappoli), caratterizzati con medie e varianze interne e, tra loro, differenziati da medie e varianze esterne.

CLUSTER E MEDIE DEI CLUSTER	VARIANZE NEI CLUSTER SQM NEI CLUSTER				DISTANZE DALLA MEDIA GENERALE (E LORO QUADRATI)	
1°	2	1	0	0	2.14	4.5630
2°	4	1.5	0.250	0.5	2.13	4.5549
3°	2.5	3.42	1.7249	1.31	0.36	0.1309
VARIANZA RESIDUA			1.4628	VARIANZA SPIEGATA		0.8452
SQM RESIDUO			1.21	SQM SPIEGATO		0.92

Per evidenziare i cluster, in modo significativo nel test di Fisher (come pure nel test non parametrico di Siegel – Tukey), è utile moltiplicare per un peso opportuno tutte le distanze e, nello specifico per 7.5, cosicché il risultato prima non ancora del tutto significativo, risulti poi significativo, sia conformando meglio il terzo cluster originale, sia calcolandone la dispersione passo a passo, anziché come varianza.

VAR. SPIEGATA “PESATA”	0.8452
SQM SPIEGATO “PESATO”	0.92

#### TEST DI FISHER

VALORE SPERIMENTALE	4.33 <sup>10</sup>	GDL	2	28
VALORI TEORICI	3.34 (ALFA = 5% <sup>11</sup> )	5.45 (ALFA = 1%)		

#### TEST DI SIEGEL – TUKEY<sup>12</sup>

DISTANZE	0	0.5	0.5	0.65	0.65	0.76	1.56	1.61	2.71	16.01	16.02
RANGHI	1	2	3	4	5	6	7	8	9	10	1
INT. / EXT.	I	I	I	I	I	I	I	I	E	E	E

SOMMA DEI RANGHI INTERNI	36	
VALORE SPERIMENTALE	2.45	
VALORI TEORICI	1.95 (ALFA = 5%)	2.57 (ALFA = 1%)

L'esito del test colloca questo problema nella zona del dubbio, dove non è possibile trarre con sicurezza una risposta. Per questa regione, la cluster analysis è qui riproposta con uno dei tre cluster, meno disperso, allo scopo di provare l'effettivo guadagno della varianza residua sulla varianza generale, dato il formarsi di una varianza spiegata significativa (essendo le nuove MEDIE GENERALI 2.84 3.35).

<sup>10</sup> Qui come nel seguito, sono sottolineati i valori sperimentali significativi, rispetto ai test eseguiti.

<sup>11</sup> Con la lettera: alfa, si indica usualmente il livello di significatività del test.

<sup>12</sup> Il test di Siegel – Tukey è un test non parametrico (in inglese, detto: *distribution free*, benché i due termini non siano sinonimi), svolto qui solo per la cluster analysis (nell'ipotesi che i dati non siano normali), allo scopo di mostrare la praticabilità anche di questi test. Nello specifico, questo test mostra che, se le varianze interne sono tutte più piccole delle varianze esterne, allora i cluster sono ben formati e sufficientemente lontani tra loro.

X / Y	1	2	3	4
1				
2	3		2	9
3			4	4
4	1	1		7

MEDIE DEI CLUSTER			VARIANZE NEI CLUSTER SQM NEI CLUSTER		DISTANZE DALLA MEDIA GENERALE (E LORO QUADRATI)	
1°	2	1	0	0	2.50	6.2487
2°	4	1.5	0.25	0.5	2.19	4.7890
3°	2.85	3.77	0.8461	0.92	0.41	0.1718
VARIANZA RESIDUA			0.7258	VARIANZA SPIEGATA		1.0578
SQM RESIDUO			0.85	SQM SPIEGATO		1.03
			VAR. SPIEGATA "PESATA"		7.9331	
			SQM SPIEGATO "PESATO"		2.82	

#### TEST DI FISHER

VALORE SPERIMENTALE	<u>10.93</u>	GDL	2	28
VALORI TEORICI	3.34 (ALFA = 5%)	5.45 (ALFA = 1%)		

Per contro, se nel terzo cluster originale, in quanto molto allungato, non si calcolano gli scarti rispetto alla media del cluster, ma le distanze, elemento ad elemento, dividendo per il numero degli elementi, aumentato di un numero opportuno (ad esempio, una volta e mezzo il numero degli elementi), in quanto tutti gli elementi (tranne il primo e l'ultimo) sono contati due volte, allora si hanno i seguenti valori:

"VARIANZA" e "SQM" DEL 3° CLUSTER	0.8521	0.92
VARIANZA e SQM RESIDUA/O	0.7308	0.85
VALORE SPERIMENTALE DEL TEST DI FISHER	<u>8.67</u>	

Infatti la varianza, come quasi tutti gli indici classici di dispersione, prendendo in considerazione insiemi a più dimensioni, ha un comportamento di tipo isotropo che penalizza gli insiemi di dati, disposti su figure sottili, in una o più dimensioni, ed allungate in una o poche altre.

Al contrario, indici che calcolano distanze fra elementi vicini (senza prendere in considerazione baricentri e centroidi), pesandole opportunamente, tengono proprio conto di questi fattori di forma <sup>13</sup>.

#### ANOVA <sup>14</sup>

L'analisi dei segni e dei valori assoluti (in questo caso specifico, dei valori assoluti, rispettivamente minori e maggiori di uno) dei dati, compresi nella lista di valori (che ha già condotto allo studio delle statistiche ottimali

<sup>13</sup> Per completezza, si osserva, come la semplicità dell'esempio dia già individuati i cluster ed i loro elementi, mentre basi di dati, più grandi, richiedano l'adozione di differenti strategie (agglomerative o divisive, oppure sequenziali), preliminarmente all'analisi statistica per la validazione dei dati, raggruppati nei cluster, ed alla ricerca di miglioramenti, per evidenziare la loro buona conformazione.

<sup>14</sup> ANOVA (ovvero Analysis of Variance) è la sigla inglese per l'Analisi di varianza.

e di quelle robuste della variabile statistica semplice), permette di individuare quattro diversi raggruppamenti di osservazioni, caratterizzati con medie e varianze interne, e differenziati da medie e varianze esterne.

#### DATI PARTIZIONATI (IN 4 RAGGRUPPAMENTI)

0.09	0.77	0.00	0.48	0.56	0.98	0.13
1.13	1.91					
-0.42	-0.44	-0.18	-0.24			
-2.01	-1.19	-1.31				
7	2	4	3	NUMERO DEI DATI		
6	1	3	2	GDL		
0.43	1.52	-0.32	-1.50	MEDIE CONDIZIONATE		
0.1183	0.1521	0.0126	0.1308	VARIANZE CONDIZIONATE		
0.34	0.39	0.11	0.36	SQM CONDIZIONATI		

Analogamente a quanto sviluppato nella cluster analysis, anche l'analisi di varianza, procede al calcolo di una varianza spiegata (come varianza delle medie condizionate, rispetto alla media generale) e poi di una varianza residua (come media delle varianze condizionate), procedendo infine al loro confronto statistico con il test di Fisher <sup>15</sup>.

VARIANZA RESIDUA	0.0984	VARIANZA SPIEGATA	0.8187
SQM RESIDUO	0.31	SQM SPIEGATO	0.90

#### TEST DI FISHER

VALORE SPERIMENTALE	<u>8.32</u>	GDL	3	12
VALORI TEORICI	3.49 (ALFA = 5%)	5.95 (ALFA = 1%)		

Dopodiché riconosciuto l'ottimo comportamento delle classi condizionate, ovvero il significativo differenziarsi delle medie condizionate, rispetto alla media generale (tenuto anche conto del valore limitato delle varianze condizionate e conseguentemente della varianza residua), altri raggruppamenti maggiori sono qui presi in considerazione, allo scopo di verificarne la significatività.

Infatti un primo raggruppamento confronta due insiemi, diversi solo per il segno dei loro elementi, mentre un secondo raggruppamento confronta altri due insieme, questa volta, diversi per il valore assoluto sempre dei loro elementi. Tuttavia nessuno dei due test dà una chiara risposta positiva e, mentre il primo si colloca nella regione nulla, il secondo mostra addirittura il prevalere della varianza residua sulla varianza spiegata.

#### PRIMO RAGGRUPPAMENTO

9	7	NUMERO DEI DATI
8	6	GDL

<sup>15</sup> L'analisi di varianza robusta è un'alternativa a quella classica e si ha calcolando mediane, invece di medie, e mav, invece di varianze; dopodiché il mav spiegato (mav delle singole mediane) sostituisce la varianza spiegata ed il mav residuo (mediana dei singoli mav) la varianza residua. In questo caso, il test di Fisher non è applicabile (data la non normalità dei dati e delle stime) e bisogna procedere con test non-parametrici per il confronto della dispersione.

0.67	-0.83	MEDIE CONDIZIONATE
0.3312	0.4062	VARIANZE CONDIZIONATE
0.58	0.64	SQM CONDIZIONATI

VARIANZA RESIDUA	0.3640	VARIANZA SPIEGATA	0.5532
SQM RESIDUO	0.60	SQM SPIEGATO	0.74

#### TEST DI FISHER

VALORE SPERIMENTALE	1.52	GDL 1	14
VALORI TEORICI	4.60 (ALFA = 5%)	8.86 (ALFA = 1%)	

#### SECONDO RAGGRUPPAMENTO

11	5	NUMERO DEI DATI
10	4	GDL
0.16	-0.29	MEDIE CONDIZIONATE
0.2100	2.3330	VARIANZE CONDIZIONATE
0.46	1.53	SQM CONDIZIONATI

VARIANZA RESIDUA	0.8735	VARIANZA SPIEGATA	0.0438
SQM RESIDUO	0.93	SQM SPIEGATO	0.21

#### TEST DI FISHER

VALORE SPERIMENTALE	0.050	GDL 1	14
VALORI TEORICI	4.60 (ALFA = 5%)	8.86 (ALFA = 1%)	

La mancata conferma della bontà di questi raggruppamenti, più grandi, è comunque istruttiva, perché mostra come non tutti i modelli siano significativi. Pertanto non sempre i modelli possono costituire una spiegazione utile, per l'analisi di basi di dati, se queste contengono, al loro interno, un rumore (diverso dalla presenza di dati anomali e/o errori grossolani), ben superiore ad altri segnali (forse contenuti).

A riguardo, mentre l'individuazione e la rimozione di dati anomali e/o errori grossolani, procedure robuste (e nel caso specifico, la regressione robusta), eventualmente corredata da test non parametrici e comunque seguita da test della normalità (detti "in avanti"), sono un'ottima strategia, l'individuazione uno o più segnali, in presenza di "rumori" sistematici o pseudosistematici eccessivi, è sempre un problema complesso<sup>16</sup>.

### REGRESSIONE MULTIPLA

La lista contenente i dati della variabile statistica semplice e la tabella, a doppia entrata, contenente i dati della variabile statistica doppia forniscono le osservazioni per il calcolo di due piccole regressioni multiple, rispettivamente cubica e lineare (ovviamente la seconda su un dominio bidimensionale), dove la semplicità dei due esempi permette la scrittura completa di dati e risultati.

<sup>16</sup> La rimozione di "rumori" sistematici o pseudosistematici eccessivi richiede un uso adeguato di filtri (passa alto o passa basso, oppure passa banda, ecc.); tuttavia una varianza (ovvero una dispersione) troppo grande di questi "rumori" sistematici o pseudosistematici eccessivi può purtroppo cancellare quantomeno i segnali più piccoli.

## REGRESSIONE CUBICA

Modello funzionale:  $y = a + bx^2 + cx + dx^3$

Modello stocastico caratterizzato da osservazioni indipendenti e di uguale precisione.

### DATI IN INGRESSO

1	-7.5	-2.01
2	-6.5	-1.31
3	-5.5	-1.19
4	-4.5	-0.44
5	-3.5	-0.42
6	-2.5	-0.24
7	-1.5	-0.18
8	-0.5	0.00
9	0.5	0.09
10	1.5	0.13
11	2.5	0.48
12	3.5	0.56
13	4.5	0.77
14	5.5	0.98
15	6.5	1.13
16	7.5	1.91

### MATRICE DISEGNO

1	56.25	-7.5	-421.875
1	42.25	-6.5	-274.625
1	30.25	-5.5	-166.375
1	20.25	-4.5	-91.125
1	12.25	-3.5	-42.875
1	6.25	-2.5	-15.625
1	2.25	-1.5	-3.375
1	0.25	-0.5	-0.125
1	0.25	0.5	0.125
1	2.25	1.5	3.375
1	6.25	2.5	15.625
1	12.25	3.5	42.875
1	20.25	4.5	91.125
1	30.25	5.5	166.375
1	42.25	6.5	274.625
1	56.25	7.5	421.875

### TERMINE NOTO

2.01
1.31
1.19
0.44
0.42
0.24
0.18
0
-0.09
-0.13
-0.48
-0.56
-0.77
-0.98
-1.13
-1.91

0.016

0.9172

0.958

MEDIA

VARIANZA GENERALE

SQM GENERALE

### MATRICE NORMALE

16	340	0	0
	12937	0	0
		340	12937
SIM.			582951.25

### TERMINE NOTO NORMALE

-0.26
9.78
-68.38
-2849.46

DETERMINANTI SUPERIORE 91392

INFERIORE<sup>17</sup>  $3.08 \times 10^7$

### MATRICE INVERSA

0.14156	-0.00372	0	0
	0.00018	0	0
		0.0189	-0.000419522
SIM.			0,0000110255528212185

### SOLUZIONE

0.07317
-0.00268
0.09725
0.00273

<sup>17</sup> Il determinante superiore è riferito al minore diagonale principale superiore (due per due) ed il determinante inferiore è riferito al minore diagonale principale inferiore (due per due), essendo identicamente nullo il minore extradiagonale (sempre due per due). Proprio questo annullamento è la ragione dell'ordinamento separato tra potenze pari (ovvero: zero e due) e potenze dispari (ovvero: uno e tre), dato un dominio simmetrico dei dati in ingresso. A riguardo, si noti invece come un'eventuale mancanza di simmetria nella base dei dati faccia sì che anche il minore extradiagonale non sia identicamente nullo (in questo caso, essendo quattro la dimensione della matrice normale, sono necessarie tecniche numeriche adeguate per il calcolo della matrice inversa).

## DATI INTERPOLATI

-1.959
-1.422
-0.997
-0.667
-0.417
-0.229
-0.088
0.024
0.121
0.222
0.342
0.498
0.705
0.981
1.342
1.804

0.9043  
0.951

## SCARTI RESIDUI

0.051
-0.112
0.193
-0.227
0.003
0.011
0.092
0.024
0.031
0.092
-0.138
-0.062
-0.065
0.001
0.212
-0.106

0.0172  
0.131

VARIANZA SPIEGATA E RESIDUA  
SQM SPIEGATO E RESIDUO

## TEST DI FISHER

VALORE SPERIMENTALE 52.664 GDL 16 12  
VALORI TEORICI 2.60 (ALFA = 5%) 3.98 (ALFA = 1%)

L'esiguità degli scarti residui spiega esaurientemente il successo della regressione cubica, dove la varianza spiegata è prossima alla varianza generale e la varianza residua è pressoché nulla. Invece se l'osservazione del settimo dato fosse inserita omettendo il punto, diciotto centesimi diventerebbero diciotto interi ed il dato sarebbe anomalo, completamente fuori dalla curva interpolata.

Allora sta alla conoscenza specifica della base di dati qui analizzata stabilire, se questo dato costituisce un errore grossolano, individuato e da eliminare, oppure un dato anomalo che, se di sicuro interesse, richiede tuttavia la costruzione di un modello funzionale più complesso (che, nel caso specifico, sembra essere un salto finito, da trattare, come tale, in modo separato, dal resto dei dati).

Pertanto ancora una volta, non si hanno modelli funzionali (così come modelli stocastici e norme) giusti o sbagliati, ma modelli funzionali (così come modelli stocastici e norme) adatti alla modellazione di determinati fenomeni, oppure poco adatti alle finalità attese (perché decidere chi e cosa è anomalo, oppure un errore, è sempre e solo una scelta ragionata e di buon senso).

A riguardo, un'osservazione interessante rileva come la matematica e la matematica applicata, acquisite certe regole (né più, né meno dell'ortografia, la grammatica e la sintassi di una lingua), siano il regno vasto del possibile e non un recinto chiuso, composto di strade obbligate. Infatti solo una conoscenza scarsa della matematica può portare ad una simile conclusione <sup>18</sup>.

<sup>18</sup> Il miglior invito conseguente è sempre quello di studiare molto e sperimentare liberamente. D'altra parte, se esistono corsi teorici di Matematica elementare da un punto di vista superiore, parafrasando questa denominazione, le applicazioni pratiche più interessanti sono quelle di Matematica superiore da un punto di vista elementare.

DATI IN INGRESSO

1	-7.5	-2.01
2	-6.5	-1.31
3	-5.5	-1.19
4	-4.5	-0.44
5	-3.5	-0.42
6	-2.5	-0.24
7	-1.5	-18
8	-0.5	0.00
9	0.5	0.09
10	1.5	0.13
11	2.5	0.48
12	3.5	0.56
13	4.5	0.77
14	5.5	0.98
15	6.5	1.13
16	7.5	1.91

MATRICE DISEGNO

1	56.25	-7.5	-421.875
1	42.25	-6.5	-274.625
1	30.25	-5.5	-166.375
1	20.25	-4.5	-91.125
1	12.25	-3.5	-42.875
1	6.25	-2.5	-15.625
1	2.25	-1.5	-3.375
1	0.25	-0.5	-0.125
1	0.25	0.5	0.125
1	2.25	1.5	3.375
1	6.25	2.5	15.625
1	12.25	3.5	42.875
1	20.25	4.5	91.125
1	30.25	5.5	166.375
1	42.25	6.5	274.625
1	56.25	7.5	421.875

TERMINE NOTO

2.01
1.31
1.19
0.44
0.42
0.24
18
0
-0.09
-0.13
-0.48
-0.56
-0.77
-0.98
-1.13
-1.91

-1.098  
19.9609  
4.468

MEDIA  
VARIANZA GENERALE  
SQM GENERALE

MATRICE NORMALE

16	340	0	0
	12937	0	0
		340	12937
SIM.			582951.25

TERMINE NOTO NORMALE

17.56
49.87
-95.11
-2909.60

SOLUZIONE

-2.30018
0.0566
0.57732
-0.00782

DATI INTERPOLATI

-0.147
-1.514
-2.462
-3.039
-3.292
-3.268
-3.012
-2.574
-1.998
-1.333
-0.625
0.078
0.731
1.286
1.69
1.914

SCARTI RESIDUI

1.863
-0.204
-1.272
-2.599
-2.872
-3.028
14.988
-2.574
-2.088
-1.463
-1.105
-0.482
-0.039
0.306
0.566
0.004

3.1531

22.4104

VARIANZA SPIEGATA E RESIDUA

1.776

4.734

SQM SPIEGATO E RESIDUO

## TEST DI FISHER

VALORE SPERIMENTALE 0.141 GDL 16 12  
 VALORI TEORICI 2.60 (ALFA = 5%) 3.98 (ALFA = 1%)

La prevalenza della varianza residua sulla varianza spiegata (da cui un valore sperimentale del test di Fisher addirittura minore di uno) è una chiarissima evidenza della presenza di un dato anomalo che, come già detto in precedenza, può essere interpretato come un errore grossolano, oppure richiedere l'adozione ad hoc di un modello funzionale più complesso.

## REGRESSIONE LINEARE 2D

Modello funzionale:  $z = a + bx + cy$   $z - \bar{z} = b(x - \bar{x}) + c(y - \bar{y})$

Modello stocastico caratterizzato da osservazioni indipendenti e di uguale precisione.

## DATI IN INGRESSO

Z = F(X,Y)

X / Y	1	2	3	4
1			9	
2	3		2	
3			4	4
4	1	1		7

X / Y	-1.625	-0.625	0.375	1.375
-1.875			5.125	
-0.875	-0.875		-1.875	
0.125			0.125	0.125
1.125	-2.875	-2.875		3.125

2.875 2.625 3.875 MEDIE X, Y, Z <sup>19</sup>  
 7.1094 VARIANZA GENERALE  
 2.666 SQM GENERALE

## MATRICE DISEGNO

## TERMINE NOTO

-0.875	-1.625
1.125	-1.625
1.125	-0.625
-1.875	0.375
-0.875	0.375
0.125	0.375
0.125	1.375
1.125	1.375

-0.875
-2.875
-2.875
5.125
-1.875
0.125
0.125
3.125

## MATRICI NORMALE ED INVERSA

## TERMINE NOTO NORMALE

## SOLUZIONE

8.875	-0.375
SIM.	9.875

0.112857	0.004286
SIM.	0.101429

10.125
-13.625

-1.0843
1.3386

<sup>19</sup> Il passaggio dai dati originari ai dati a media nulla permette di togliere dal sistema il termine noto dell'equazione di osservazione (nel caso specifico, di un piano). A posteriori, non è poi difficile sommare ai dati interpolati la loro media, ottenendo così dati interpolati dello stesso tipo di quelli originari.

DATI INTERPOLATI

2.649
0.480
1.819
6.410
5.326
4.241
5.580
4.496

3.6520  
1.911

SCARTI RESIDUI

-0.351
-0.520
0.819
-2.590
3.326
0.241
1.580
-2.504

4.6098  
2.147

VARIANZA SPIEGATA E RESIDUA  
SQM SPIEGATO E RESIDUO

TEST DI FISHER

VALORE SPERIMENTALE 0.792 GDL 8 6  
VALORI TEORICI 4.15 (ALFA = 5%) 8.10 (ALFA = 1%)

Questa volta, la prevalenza della varianza residua sulla varianza spiegata (da cui un valore sperimentale del test di Fisher addirittura minore di uno) mette in evidenza la presenza di un dato anomalo (con una sezione a V: 9, 2, 4, 7, ovviamente non-lineare).che non necessariamente deve essere interpretato come un errore grossolano, ma richiederebbe invece l'adozione di un modello funzionale più complesso <sup>20</sup>.

DATI IN INGRESSO

$$Z = F(X, Y)$$

X / Y	1	2	3	4
1			9	
2	3		7	
3			4	4
4	1	1		2

X / Y	-1.625	-0.625	0.375	1.375
-1.875			5.125	
-0.875	-0.875		3.125	
0.125			0.125	0.125
1.125	-2.875	-2.875		-1.875

TERMINE NOTO

TERMINE NOTO NORMALE

SOLUZIONE

DATI INTERPOLATI

SCARTI RESIDUI

-0.875
-2.875
-2.875
5.125
3.125
0.125
0.125
-1.875

20.125
-8.625

-2.2343
0.7886

4.549
0.080
0.869
8.360
6.126
3.891
4.680
2.446

1.549
-0.920
-0.131
-0.640
-0.874
-0-109
0.680
0.446

<sup>20</sup> Ad esempio, un paraboloide iperbolico (od a sella, cioè una funzione bilineare) potrebbe essere un modello funzionale più adeguato.

VARIANZA SPIEGATA E RESIDUA	6.4708	0.8514
SQM SPIEGATO E RESIDUO	2.544	0.923

### TEST DI FISHER

VALORE SPERIMENTALE	7.600	GDL 8	6
VALORI TEORICI	4.15 (ALFA = 5%)	8.10 (ALFA = 1%)	

In questo caso, un semplice scambio fra i valori delle quote 2 e 9 fa sì che il test di Fisher (con una varianza spiegata maggiore della varianza residua) si collochi già nella regione del dubbio. Dopodiché riconoscendo un migliore andamento lineare, se si abbassa a 0 il valore della quota pari ad 1 (più a sinistra) e si innalza a 4 il valore della quota pari a 3, il test di Fisher diventa ampiamente significativo <sup>21</sup>.

### DATI IN INGRESSO

$Z = F(X, Y)$

X / Y	1	2	3	4
1			9	
2	4		7	
3			4	4
4	0	1		2

X / Y	-1.625	-0.625	0.375	1.375
-1.875			5.125	
-0.875	0.125		3.125	
0.125			0.125	0.125
1.125	-3.875	-2.875		-1.875

### TERMINE NOTO

#### TERMINE NOTO NORMALE

#### SOLUZIONE

#### DATI INTERPOLATI

#### SCARTI RESIDUI

0.125
-3.875
-2.875
5.125
3.125
0.125
0.125
-1.875

22.125
-8.625

-2.4600
0.7800

4.760
-0.160
0.620
8.780
6.320
3.860
4.640
2.180

0.760
-0.160
-0.380
-0.220
-0.680
-0.140
0.640
0.180

VARIANZA SPIEGATA E RESIDUA	7.6444	0.2867
SQM SPIEGATO E RESIDUO	2.765	0.535

### TEST DI FISHER

VALORE SPERIMENTALE	26.666	GDL 8	6
VALORI TEORICI	4.15 (ALFA = 5%)	8.10 (ALFA = 1%)	<sup>22</sup>

<sup>21</sup> Si badi bene che questa ulteriore modifica della base dei dati non significa affatto camuffare i dati (cosa che non deve assolutamente essere fatta mai), ma semplicemente mettere in evidenza come dati diversi diano risultati differenti.

<sup>22</sup> Esempi più complessi, per numero dei dati, dimensioni dei problemi, modelli adottati (funzionali e stocastici), ecc., sono sicuramente molto più significativi, dal punto di vista dell'importanza di esempi reali, ma presentano evidenti difficoltà, per un'esposizione completa (data la loro estensione), riducendo così notevolmente il significato degli esempi stessi.

### Ibridare la Geomatica

Le proposte metodologiche, per costruire programmi di ricerca scientifica e tecnologica, non rispondono a domande “ultime”, quasi a voler cercare una fittizia unità, a partire da un’origine comune, artificialmente inventata, ma a domande “penultime” che sappiano riconoscere l’ampia collezione delle modalità e dei fini. Tutto ciò è vero nei periodi di scienza normale ed in quelli di rivoluzione scientifica, eppure forse i primi sono più interessanti, perché narrano la ricerca di ogni giorno e non alcune rare eccezioni. A riguardo, proposte e programmi (come pure resoconti ed lavori successivi) non dovrebbero mai essere vuoti elenchi, per revisori ciechi ed estranei, ma basi per discussioni serie che, partendo dai più vicini, si diffondano, per via osmotica in contesti via, via più ampi, favorendo il confronto, la critica e la crescita.

In questo modo, invece di cercare un inesistente principio regolatore di ordine superiore, capace di presiedere a tutto e per tutti, occorre fare passi concreti, a partire dalla varietà dei campi del sapere e dell’agire, verso utili meccanismi autogestiti e poi anche di cogestione, libera ed antiautoritaria. A riguardo, è necessario fare riferimento alla prassi usuale di ricerca ed azione, per riconoscere premesse culturali, analisi scientifiche e modalità tecnologiche, ed i loro effetti in altri campi della ricerca e dell’azione (specificando se misurabili e se in modo quantitativo). Infatti nel vasto campo della Geomatica e della Geomatica Applicata, a partire dalla loro nobile origine nell’Astronomia e nella Geodesia, per diffondersi poi in tutte le discipline del rilevamento, misure, modelli (quantitativi) e calcoli sono una caratteristica precipua.

Tutto ciò è certamente lontano dal trascendentalismo (religioso, oppure ateo), così come dall’idealismo ed ancora dallo storicismo. D’altra parte, queste severe domande fanno riferimento al concetto di somiglianze di famiglia, caratterizzate dall’incertezza e dal dubbio. Inoltre queste stesse domande si collegano a tradizioni rigorose e radicali, dove una lunga comunanza storica, fatta addirittura dalle stesse persone, unisce tutte le scienze geodetiche e cartografiche, con le loro scienze affini e le tecniche applicate. La stessa unione si ha, in primis, con la matematica, la fisica e le altre scienze della natura, ma poi anche con le scienze umane, le lettere e le arti che, tutte insieme, formano quel contesto culturale, caratterizzante le società più evolute, di ogni tempo storico ed ogni contesto geografico.

Infatti le dinamiche dei linguaggi e, soprattutto oggi (con la loro fortissima smaterializzazione) tutte le discipline del rilevamento sono soprattutto un linguaggio, potrebbero / dovrebbero permettere interpretazioni più generali di quelle tradizionali, fondate sul riduzionismo. Di conseguenza, muovendosi da costellazioni di culture e discipline, anche molto lontane e fortemente dissimili, un punto d’arrivo significativo costruisce una galassia di alternative, capaci di ibridare, organizzandoli armoniosamente, i vari campi del sapere. Parole ed oggetti sarebbero allora parti di una verità, ancora nascosta, certamente precaria e sicuramente provvisoria, tuttavia richiesta per compiere, ciascuno e tutti insieme, quei passi necessari, a tutela, mitigazione e crescita, del mondo e nella/e sua/e società umana/e.

Una precisazione doverosa riguarda qui il singolare del mondo, sostanzialmente unico da un punto di vista fisico e geografico, soprattutto di fronte alle odierne capacità umane di alterare ed addirittura sconvolgere i suoi meccanismi naturali. Invece una seconda precisazione doverosa riguarda il singolare / plurale riguardo la/e società umana/e. Infatti mentre è sicuramente unica la specie umana ed è razzista qualsiasi pensiero

---

<sup>23</sup> Questa prima parte cerca di dare forma di testo ad appunti, liberamente presi da chi scrive, durante il Convegno interdisciplinare Analisi: storia, teoria e pratica di un metodo tra Matematica e Filosofia (Trento, 8 – 0 giugno 2015).

contrario, sono tante le culture umane, verso le quali sono ben necessarie rispetto, tolleranza e conoscenza reciproca, essendo il multiculturalismo, il meticcio ed il sincretismo l'importante punto d'arrivo per un vero e proprio politeismo culturale. Pertanto contro ogni guerra di religione ed ogni scontro di civiltà, piccoli passi umili sono quello che si richiede soprattutto qui ed oggi.

Di nuovo, al tema qui centrale della Geomatica e della Geomatica Applicata, così come ad altre scienze e tecniche della natura, attività umane come raccogliere dati, misurare e costruire modelli sono considerate, troppo spesso, cose ovvie e naturali, senza alcuna necessità di essere messe in discussione. Infatti stabilito convenzionalmente alcuni simboli, raggiunto qualche accordo sulle procedure ed osservate le regole logiche per la conservazione della verità, una lista di numeri od una tabella, a doppia entrata, di dati potrebbero sembrare qualcosa di oggettivo. Invece la lunghissima storia della scienza e della tecnica mette in evidenza che, non sempre e non dappertutto, si sono raccolti dati, si è misurato e si sono costruiti modelli <sup>24</sup>, per ottenere mappe della conoscenza.

Pertanto proprio rifacendosi a questa lunghissima storia, si può evidenziare quali siano le varie condizioni, sia naturali che culturali, per l'emergere di queste metodologie e procedure, e come si siano formate queste specifiche discipline, a partire da questi primi concetti. Allo stesso modo, si può poi descrivere tutta la loro tradizione, inizialmente pratica ed operativa, fino al loro consolidamento e successivo inquadramento in un apparato normativo, fondante discipline, non lontane tra loro, fra loro collegate ed in mutuo divenire (così come esse sono guidate da una, più generale, ricerca scientifica e da molteplici ritrovamenti ed innovazioni tecnologici/che). Allora l'importanza, qui ed oggi, di queste discipline è dimostrata dalla grandissima varietà degli strumenti di misura, trattamento, analisi, archiviazione e visualizzazione impiegati.

Quanto segue, vuole dare impulso ad una riflessione, di vasta portata, sull'apertura di tutte le discipline del rilevamento, rispetto ad un contesto culturale molto più ampio, e promuovere, d'altro canto, un approccio che mostri quanto tali discipline abbiano contribuito a plasmare quei concetti di dato, numero e misura da cui sembrano dipendere. In questo modo, le suddette parole chiave di queste discipline: misura, trattamento, analisi, archiviazione e visualizzazione, trovano una loro centralità nella parola: analisi, perché proprio alla base di dato, numero e misura, così come di misura, modello e calcolo, che sono tipici delle discipline del rilevamento, oggi, già in anni passati, antichissimi o più prossimi, e certamente anche in anni futuri, per quanto lontani.

### **Un lungo cammino / discorso**

La geometria e l'analisi matematica nelle valutazioni di Euclide, non disgiunte dall'astronomia di alcuni suoi contemporanei, per quanto riguarda il mondo antico, ed il passaggio dalla geometria classica alla matematica moderna, sul crinale tra il tardo Rinascimento e la prima metà del Secolo d'oro, sono tappe obbligate, per conoscere come si avvia e riprende questo lunghissimo cammino / discorso, tenuto conto che, se l'ateniese Platone fonda l'Accademia delle idee (dove studiare e riflettere), il barbaro Aristotele fonda il Liceo peripatetico (dove conoscere e sperimentare).

Dopodiché il Secolo d'oro e gli albori della successiva Età dei lumi, caratterizzati dall'andare (ed è ancora un cammino, anche contro le chiusure ideologiche delle guerre di religione) verso un modello di scienza certa, sono presi in considerazione prima con l'analisi ed il *regressus* nella versione francese delle Meditazioni di Cartesio e poi andando all'origine dell'analisi come teoria matematica, così come si passa dalla velocità alle

---

<sup>24</sup> Antichissima è comunque la rappresentazione di luoghi e situazioni (e la fotogrammetria, con la camera lucida, precede la fotografia).

flussioni nella meccanica di Newton, nonché con l'analisi dei concetti e di relazioni tutto – parte nel programma logico di Leibniz.

Infine andando oltre la sintesi critica kantiana ed arrivando all'oggi, uno sguardo è gettato sulla matematica moderna e, in particolare, sulla sua analiticità e sugli assiomi della teoria degli insiemi. In parallelo, il metodo analitico in metafisica (in conformità al suo significato letterale, cioè andando oltre le cose fisiche, dove il discorso è ancora cammino e viceversa) ed alcuni elementi dell'identità della filosofia analitica sono trattati, relativamente alla filosofia contemporanea. Rimarchevole e di notevole interesse tutt'oggi, è voler mettere insieme linguaggi ed esperienze di discipline diverse.

### Da Aristotele a Cartesio

Già Platone, come riferito da Aristotele, distingue l'analisi dalla sintesi. Infatti:

- ❑ l'analisi è a priori ed intende risalire dagli effetti alle cause, cosicché assunto uno o più cercato/i ignoto/i, si dimostra la verità di fatti e/o cose noti/e;
- ❑ la sintesi è a posteriori e studia gli effetti, in base all'esperienza acquisita, a partire dalle loro varie cause, come sarà poi descritto e sistematizzato nella sintesi critica kantiana.

A riguardo, una chiara giustapposizione<sup>25</sup> è formulata da Erone di Alessandria che la riporta sintetizzandola dagli Elementi di Euclide. Ovviamente almeno in una logica moderna, scettica e relativista moderata, occorre precisare che l'ignoto riguarda solo l'uomo e non un qualche dio che tutto sa (ma che poi forse non esiste nemmeno, tranne pensarlo per soccorrere le umanissime ansia ed angoscia).

Nella sintesi, lo strumento principe per un ragionamento valido e rigoroso è il sillogismo. Tuttavia oltre alla concatenazione logica, se le preposizioni sono convertibili, tra soggetti e predicati, allora le inferenze, da premesse a conclusioni, sono valide (ed altrimenti no). Di seguito, due esempi provano la convertibilità ed il suo contrario<sup>26</sup>:

- ❑ di giorno c'è luce, ma non c'è luce, allora non è giorno (con l'eccezione rara di un'eclisse solare);
- ❑ in inverno può nevicare, ma se nevicava non è detto che sia inverno (anche dopo una lunga statistica).

In questo modo, in matematica, si ha necessità, cioè sintesi, per dimostrare teoremi, ed analisi, per eseguire algoritmi (che non dimostrano, ma possono utilmente risolvere). Invece i processi naturali richiedono solo contingenza (notando che, nel mondo antico, la fisica e l'astronomia appartengono alla matematica, mentre i processi naturali riguardano le scienze naturali) e le azioni umane la pura libertà o quasi.

I pochi decenni compresi tra il tardo Rinascimento e la prima metà del Secolo d'oro segnano il trapasso dalla riscoperta europea del mondo antico, anche nelle sue componenti greche ed ellenistiche, andate perse nei secoli bui dell'alto medioevo, e la formazione della Scienza nuova che ancora caratterizza il mondo attuale

---

<sup>25</sup> Dopo Cartesio, è invece possibile riscontrare un certo mescolamento tra i concetti di analisi e sintesi, proprio per l'ambiguità sul tema: ignoto a chi.

<sup>26</sup> Per contro, ogni singola frase, al più, può essere poco usuale (ovvero: ritirata dal traffico, in gergo wittgensteiniano), ad esempio: l'apparecchio elettrico od elettronico dorme (invece di è spento) che tuttavia può andare bene in una storia fantastica di oggetti animati, per quanto rara. Per questa stessa ragione, anche un sillogismo può essere formalmente corretto, ma privo di senso, come ad esempio: la pesca è un frutto che matura in estate, la pesca è un modo di produzione e/o un'attività sportiva (giocando sull'uguale scrittura di una diversa modulazione della vocale: e), essendo tuttavia assurda la conclusione, formalmente ineccepibile: un frutto che matura in estate è un modo di produzione e/o un'attività sportiva.

(in un mix complesso e talvolta caotico di scienza e tecnica). Infatti nei pochi anni compresi tra il 1585 ed il 1587:

- ❑ Federico Commandino (umanista, matematico e medico) traduce Pappo di Alessandria;
- ❑ Raffaele Bombelli (matematico, ingegnere ed architetto) inventa i numeri immaginari e complessi;
- ❑ Simon Stevin di Bruges (italianizzato in Simone Stevino, ingegnere, fisico e matematico fiammingo), per la prima volta, utilizza la notazione decimale <sup>27</sup>.

Dopodiché nel 1687, Cartesio pubblica il Discorso sul metodo, con la presentazione della geometria analitica <sup>28</sup>. A riguardo, è interessante rimarcare come Cartesio dia una prima definizione di curva geometrica, quale linea descrivibile con un'equazione (o con le sue classi d'equivalenza, adottando una versione più moderna della matematica).

Dopodiché sulla scorta della soluzione di un'equazione di terzo grado, ad opera di Regiomontano (pseudonimo di Johannes Müller da Königsberg, umanista, matematico ed astronomo tedesco), come il completamento di un cubo, Cartesio propone la soluzione delle equazioni di grado superiore al quarto, come l'intersezione di una parabola di grado opportunamente alto, con una circonferenza.

Ancora di Cartesio è poi la precisazione che, quando da alcuni casi particolari si vuole passare a classi più generali, occorre sempre precisare bene tutti i limiti di tali classi. Un esempio è dato dalla determinazione del centro di figure solide qualsiasi e non solo di quelle monotone (ovvero di figure scomponibili in solidi tutti decrescenti, andando dal basso verso l'alto).

## **Newton e Leibniz**

La definizione di analisi ha una sua mutazione nel tempo, passando attraverso diversi significati:

- ❑ nella scienza greca, latina ed araba, quantità incognite, trattate come note;
- ❑ dopo François Viète (matematico ed uomo politico francese) e poi per tutto il '600, famiglia di tecniche matematiche (e sinonimo di algebra);
- ❑ dopo Newton e nel '700, famiglia di teorie matematiche generali, ad esempio, come le funzioni (da cui, nel tardo '700 e poi per tutto l'800, anche analisi infinitesimale).

In questo contesto, fondamentale è il contributo di Newton che, sia in matematica che in fisica (nella meccanica, come nell'ottica) ed in astronomia, evidenzia:

- ❑ il passaggio dalla geometria alla cinematica, dove il moto è descritto dalla velocità, ed
- ❑ il successivo passaggio dalla velocità alle flussioni, intese come trasformazione di funzioni,

cosicché, superando il concetto scolastico (medioevale), secondo cui le quantità intensive caratterizzano le variazioni, una variazione definisce una quantità astratta. Allora le flussioni sono le derivate e forniscono, in

---

<sup>27</sup> La notazione attuale è tuttavia dovuta a John Napier (italianizzato in Giovanni Nepero, matematico, astronomo e fisico scozzese).

<sup>28</sup> La prima formulazione della geometria analitica è di Apollonio di Perga, ma la sua riscoperta moderna è largamente indipendente. Diversamente si ha con la prima formulazione della trigonometria di Ipparco di Nicea; infatti la riscoperta di Regiomontano dipende direttamente dalla traduzione dell'Almagesto di Claudio Tolomeo, effettuata dallo stesso.

modo puntuale, le velocità del moto, così come il loro disegno geometrico le interpreta come le tangenti, in ogni punto, alle curve geometriche che descrivono il moto stesso.

Pertanto anche il tempo diventa solo un parametro lineare, uniforme ed omogeneo, ottenuto per astrazione e generalizzazione, grazie alla potenza del nuovo metodo impiegato. Tutto ciò comporta una rifondazione della fisica e principalmente della dinamica (della meccanica), facendola passare da una scienza descrittiva, di moti vari ed urti, ad una scienza matematica, propriamente detta.

Leibniz parte dall'analisi delle relazioni tutto – parte, dove una sintesi possibile consiste nelle disposizioni e/o combinazioni (le cui espressioni sono riportate nel seguito) di quelle stesse parti, per arrivare alle dimostrazioni ed a nuovi concetti. Andando poi avanti così, fino ad ottenere nuovi linguaggi ed altre logiche, Leibniz estende questo tipo di analisi (ovvero tutto, cioè genere, e parte, cioè specie) dagli oggetti materiali ai concetti astratti.

Permutazioni:  $P_n = n!$

Disposizioni con ripetizioni:  $D_{n,k} = n^k$

Disposizioni:  $D_{n,k} = \frac{n!}{(n-k)!}$

Combinazioni:  $C_{n,k} = \binom{n}{k} = \frac{D_{n,k}}{P_k} = \frac{n!}{(n-k)! k!}$

Infatti i concetti analitici non sono solo necessari, ma anche contingenti, ovvero necessari nella totalità dei mondi possibili, come Leibniz accenna, senza tuttavia alcuna pretesa sistematica, nella Monadologia (opera oltretutto pubblicata postuma). In questo contesto, ormai prossimo alla logica di Port Royal, Leibniz privilegia l'intensione e la comprensione, proprie degli uomini, all'estensione, del regno degli animali <sup>29</sup>:

- verità intensionale: i numeri pari sono divisibili per due, senza resto alcuno;
- verità estensionale: i numeri pari sono zero, due, quattro, otto, ecc.

Di conseguenza, dimostrare verità significa analizzare due termini, fino a trovarli uguali o contenuto uno nell'altro, in quanto le parti sono nel tutto, anche se altre caratteristiche non appartengono al tutto (ad esempio, i segmenti della retta sono parti proprie della retta, ma i punti della retta non sono sue parti proprie, perché non omogenei <sup>30</sup>).

#### Classificazione tassonomica <sup>31</sup>

- Dominio
- Regno

<sup>29</sup> La logica leibnizziana anticipa quasi la logica di George Boole ed Augustus De Morgan (logici e matematici inglesi).

<sup>30</sup> Questa affermazione differisce dai concetti di retta punteggiata, piano punteggiato o rigato e spazio punteggiato o rigato, oppure costituito da piani, propri della geometria proiettiva, ma formati solo tra la fine del '700 e l'inizio dell'800, ad opera di Gaspard Monge (matematico e disegnatore francese) e Jean-Victor Poncelet (matematico ed ingegnere francese).

<sup>31</sup> La classificazione tassonomica ([https://it.wikipedia.org/wiki/Classificazione\\_scientifica](https://it.wikipedia.org/wiki/Classificazione_scientifica)) è stabilita, sia in zoologia che in botanica, da Carl von Linné, già Carl Nilsson Linnaeus (italianizzato in Linneo).

- Phylum<sup>32</sup>
  - Sottoregno
  - Superphylum
  - Subphylum<sup>33</sup>
    - Infraphylum
    - Superclasse
- Classe
  - Sottoclasse
  - Infraclasse
  - Superordine
- Ordine
  - Sottordine
  - Infraordine
  - Parvordine
- Famiglia
  - Sottofamiglia
  - Tribù
  - Sottotribù
  - Superfamiglia
  - Infratribù
- Genere
  - Sottogenere
- Specie
  - Sottospecie
  - Forma o Varietà

#### La logica di Boole e De Morgan<sup>34</sup>

L'algebra di Boole è una particolare algebra della matematica che opera sui suoi soli valori: zero ed uno, con due sole operazioni: somma e prodotto (altrimenti dette: or e and). Questa algebra gode un gran numero di proprietà (commutativa, associativa, di assorbimento, distributiva, di idem-potenza, esistenza del minimo e del massimo, esistenza del complemento), come di seguito riportato.

- |  |                             |   |                             |
|--|-----------------------------|---|-----------------------------|
| <input type="checkbox"/> commutativa:  | $a + b = b + a$             | e | $ab = ba$                   |
| <input type="checkbox"/> associativa:  | $a + (b + c) = (a + b) + c$ | e | $a(bc) = (ab)c$             |
| <input type="checkbox"/> assorbimento: | $a + (bc) = a$              | e | $a(b + c) = a$              |
| <input type="checkbox"/> distributiva: | $a(b + c) = (ab) + (ac)$    | e | $a + (bc) = (a + b)(a + c)$ |
| <input type="checkbox"/> idem-potenza: | $a + a = a$                 | e | $aa = a$                    |

<sup>32</sup> Tipo o Divisione.

<sup>33</sup> Sottotipo o Sottodivisione.

<sup>34</sup> Algebra e teoremi ([https://it.wikipedia.org/wiki/Algebra\\_di\\_Boole](https://it.wikipedia.org/wiki/Algebra_di_Boole) e [https://it.wikipedia.org/wiki/Teoremi\\_di\\_De\\_Morgan](https://it.wikipedia.org/wiki/Teoremi_di_De_Morgan)) costituiscono le basi matematiche della logica moderna.

- esistenza di minimo e massimo:  $a \times 0 = 0$  e  $a + 1 = 1$
- esistenza del complemento:  $a(\text{not } a) = 0$  e  $a + (\text{not } a) = 1$

Una legge di dualità<sup>35</sup> lega i valori: zero ed uno (essendo: zero il duale di uno e viceversa), le operazioni: somma e prodotto (essendo: la somma duale del prodotto e viceversa), ed una qualsiasi variabile ( $a$ ) con la sua negazione ( $\text{not } a$ ). Pertanto proprio la legge di dualità e l'esistenza di minimo e massimo, fanno sì che l'algebra di Boole sia dotata degli elementi neutri della somma e del prodotto, rispettivamente: zero ed uno:

$$a + (a \times 0) = a + 0 = a \qquad \text{e} \qquad a(a + 1) = a \times 1 = a \quad ^{36}$$

Conformemente alla logica aristotelica, l'operazione di convoluzione, ovvero la doppia negazione, individua l'elemento dato (cioè:  $\text{not}(\text{not } a) = a$ ). Inoltre il secondo teorema dell'assorbimento e la legge di dualità permettono l'assorbimento del complemento, sia della somma che del prodotto (cioè:  $a + (\text{not } a)b = a + b$  e  $(\text{not } a)(a + (\text{not } b)) = (\text{not } a)(\text{not } b)$ ).

Il teorema dell'elemento unico afferma poi che, se si ha:  $x + y = a$  e  $xy = 0$ , allora l'elemento  $y$  è unico, potendosi dimostrare che esiste un solo elemento, soddisfacente entrambe le espressioni. Di conseguenza, contrariamente all'aritmetica ed all'algebra classica, vale solo un doppio principio di eliminazione, ovvero se e solo se si ha insieme:  $x + y = x + z$  e  $xy = xz \Rightarrow y = z$ , cioè un risultato valido.

Infine presi due qualsiasi elementi nell'algebra di Boole, valgono anche i due teoremi di De Morgan:

$$\text{not}(ab) = \text{not}(a) + \text{not}(b) \qquad \text{not}(a + b) = \text{not}(a)\text{not}(b)$$

Per completezza, si elencano sette operatori, facilitando il trattamento delle informazioni, con questa logica:

- *not* : operatore che fornisce, in uscita, il valore inverso, rispetto a quello in entrata (a riguardo, si chiama: *buffer*, un non-operatore che lascia inalterati i valori);
- *and* : operatore che assegna, valore uguale ad uno, se tutti gli elementi, in entrata, sono uguali ad uno, ed altrimenti il valore: zero;
- *or* : operatore che assegna, valore uguale ad uno, se almeno un elemento, in entrata, è uguale ad uno, ed altrimenti il valore: zero;
- *xor*<sup>37</sup>: operatore che assegna, valore uguale ad uno, se è dispari il numero di elementi, in entrata, uguali ad uno, ed altrimenti il valore: zero;
- *nand* : operatore che assegna, valore uguale ad zero, se tutti gli elementi, in entrata, sono uguali ad uno, ed altrimenti il valore: uno (essendo semplicemente:  $\text{not}(\text{and})$ );

<sup>35</sup> Le leggi di dualità hanno una validità più generale e sono ulteriormente generalizzate dalle leggi di sostituibilità, dove diverse classi di enti matematici possono essere sostituiti da altri enti matematici che godono delle stesse proprietà formali.

<sup>36</sup> Qui, come già prima ed anche oltre, il segno: per, della moltiplicazione è scritto solo dove facilita la lettura.

<sup>37</sup> Questo operatore è altresì detto: operatore di disparità, e la sua negazione (cioè l'ultimo operatore dell'elenco): operatore di parità.

- ❑ *nor* : operatore che assegna, valore uguale ad uno, se tutti gli elementi, in entrata, sono uguale ad zero, ed altrimenti il valore: (uno essendo analogamente:  $not(or)$ );
- ❑ *xnor* : operatore che assegna, valore uguale ad uno, se è pari il numero di elementi, in entrata, uguali fra loro, ed altrimenti il valore: zero (essendo ancora:  $not(xor)$ ).

In questo modo, dati due elementi i cui valori sono, nell'ordine: (00), (01), (10) e (11), gli operatori: *and*, *or*, *nand* e *nor*, forniscono quattro risposte distinte che coprono alcune possibili combinazioni dei dati: (0001), (0111), (1110) e (1000), mentre gli operatori: *xor* e *xnor*, forniscono altre due risposte che coprono altre combinazioni ragionevoli degli stessi dati: (0110) e (1001).

### Il contributo dell'analisi matematica

L'analisi matematica affina e puntualizza i problemi posti, cosicché un problema si dice ben posto, se:

- ❑ i dati sono compatibili con le soluzioni attese;
- ❑ gli stessi dati sono sufficienti per avere un'unica soluzione;
- ❑ si sa come la soluzione dipenda da dati e dai loro errori (cosicché una buon legame, tra variazioni nella soluzione ed errori, faccia sì che piccole variazioni nei dati comportino piccole variazioni nella soluzione).

Dopodiché ogni problema richiede di sapere come:

- ❑ trovare la sua soluzione;
- ❑ stimarla opportunamente, potendosi misurare la distanza tra una soluzione teorica e l'approssimazione trovata (ad esempio, in uno spazio di Hilbert<sup>38</sup>);
- ❑ stabilirne la bontà, cioè se è vicina, rispetta le informazioni disponibili ed è praticamente computabile.

A riguardo, occorre precisare come ogni approssimazione possa avere anche solo:

- ❑ una validità locale;
- ❑ una validità in media,
- ❑ soluzioni multiple non tutte fisicamente accettabili, perché non realistiche.

In generale, l'analisi matematica si contrappone alla geometria ed all'algebra, passando rispettivamente dalla descrizione dei problemi alla loro soluzione, come pure affrontando problemi in un ambito più vasto, rispetto alle quantità discrete algebriche. Per contro, in matematica, non è così facile definire cosa significhi sintesi.

Forse in matematica, potrebbe significare provvedere ad una raccolta di esempi / applicazioni, collegandoli ad un'analogia raccolta di metodi / teorie, tutti opportunamente raggruppati e classificati, in una grandissima

---

<sup>38</sup> Lo spazio di Hilbert è uno spazio lineare, dotato di norma e con prodotto scalare (comunque anche per gli spazi solo metrici, formati da insiemi, si devono definire le dimensioni topologiche, la misura integrale e le curvature).

matrice inventario (od in qualche altro ente matematico, a più dimensioni, se necessario allo scopo), con lo scopo di documentare varie e diverse richieste con vari e diversi modi di procedere.

Infatti Kurt Gödel (matematico, logico e filosofo austriaco) definisce due tipi di analisi, distinguendo l'analisi convenzionale tautologica dall'analisi come elenco / descrizione di elementi basilari <sup>39</sup>, indipendenti e veri a priori (perché evidenti). Di conseguenza, si ha incompletezza e/o indeterminazione dei concetti matematici, solo se essi sono privi di valore, perché non analiticamente veri, né falsi.

In questo contesto, la costruzione di una gerarchia degli insiemi, secondo un metodo di massimizzazione del risultato atteso, porta ad includere negli insiemi superiori anche quegli insiemi dipendenti dall'assioma della scelta. Tutto ciò prova ulteriormente come la matematica e, in particolare, la matematica moderna sia un campo aperto alla ricerca libera e non una collezione arida di verità rivelate assolute.

### **La filosofia analitica**

La filosofia analitica deriva dalla logica di Friedrich Ludwig Gottlob Frege (matematico, logico e filosofo tedesco) e Bertrand Arthur William Russell (matematico, logico, filosofo ed attivista politico inglese), come metodo e da un punto di vista ontologico, fino ad arrivare ad una conciliazione tra la metafisica aristotelica e la sintesi critica kantiana.

Infatti oggi, dismessi i panni desueti ed i sogni regressivi della religione (qualunque essa sia) e delle ideologie, la metafisica si occupa della struttura profonda fondamentale della realtà, cioè di ciò che esiste e può esistere, secondo la proposta di Platone nel Fedro che consiste nel definire la migliore scomposizione, cioè i tagli più opportuni, per l'individuazione degli elementi semplici nell'essere.

La nuova metafisica parte dalla critica di Ludwig Josef Johann Wittgenstein (ingegnere, logico e filosofo austriaco), Rudolf Carnap (filosofo e logico tedesco) ed Alfred Tarski (matematico, logico e filosofo polacco), contro la metafisica classica, perché priva di logica e, proprio in base alle critiche fatte dalla logica stessa, è descrittiva e realista, in accordo con la metafisica kantiana.

La logica del linguaggio opera in modo neutro, non dogmatico e leggero, trattando proposizioni, predicati <sup>40</sup>, qualificatori <sup>41</sup> e mondi possibili. Dopodiché le funzioni sono entità insature, che sono saturate da argomenti / oggetti (con i loro attributi o meno), per formare preposizioni, a loro volta, vere o false (oppure anche solo probabili, in una logica a tre valori), secondo le somiglianze di famiglia.

I mondi possibili sono descritti semanticamente, dove la semantica utilizzata mette in gioco una metafisica leggera (data a priori) e chi li popola, compreso chi giudica, è associato in base a criteri di somiglianza. In questo contesto, il mescolamento, già cartesiano, tra i concetti di analisi e sintesi, si approfondisce ancora, per una sempre maggiore relatività dei punti di vista.

In tempi ancora più recenti, la controversia tra la filosofia analitica e l'analisi porta al superamento del loro legame (come nella teoria politica di John Bordley Rawls, filosofo politico americano), al di là dell'uso di un metodo rigoroso. Tuttavia oltre all'importanza del contributo storico, due concetti dell'analisi svolgono tuttora un ruolo importante nella filosofia analitica:

□ il patrimonio storico;

---

<sup>39</sup> Questa formulazione gödeliana non è troppo lontana dalla definizione, di Kant, sulla matematica come sintetico a priori. Da'altra parte, nelle sue ultime formalizzazioni metodologiche, Gödel si avvicina proprio alla fenomenologia di Husserl.

<sup>40</sup> I predicati base sono: la congiunzione (e), la disgiunzione (o) e la negazione (non), dove il verbo essere è l'unico verbo usato.

<sup>41</sup> I qualificatori base sono: per ogni ed esiste (cioè per qualche).

□ la cassetta degli attrezzi (ad esempio, la parafrasi e l'analisi concettuale).

Infatti le parafrasi trasformano enunciati problematici in altri enunciati non più problematici (ad esempio, certi problemi geometrici sono trasformati in problemi algebrici, rendendoli più trattabili, di gran lunga). Tuttavia le parafrasi non sono in grado di decidere, dipendendo sempre dal linguaggio usato, cosa che rende circolari i problemi da risolvere.

Invece l'analisi concettuale è connettiva e tratta i concetti non riducendo i concetti a concetti più semplici, ma in base a concetti più chiari, nel loro significato, come proposto da Willard Van Orman Quine (filosofo e logico americano). In questo modo, un linguaggio migliore (ad esempio, come sono tutti i linguaggi naturali <sup>42</sup>) può certamente esistere e, a sua volta, rendere possibile la formazione rigorosa delle parafrasi.

Infatti un certo linguaggio target conserva le proprietà inferenziali del linguaggio naturale, ma anche altri linguaggi target possono conservare le stesse proprietà, previa l'adozione di diverse traduzioni, ma tenendo conto della distinzione tra quello che si dice e ciò che significa (cosa che attualmente rende diversi gli uomini dalla macchine, come attestato da John Rogers Searle (filosofo analitico americano), così:

le credenze di quei signori anziani sono un po' antiquate (liberamente ispirata a Donald Davidson, filosofo analitico americano): si riferisce ai pensierini od agli armadietti,

e cose visibili possono essere invisibili. Se qualcuno va a cavallo in un bosco, prima lo si vede, poi no, ma si sa che c'è. Nella Firma in bianco, la cavallerizza nasconde gli alberi e gli alberi la nascondono a loro volta. Tuttavia il nostro pensiero comprende tutti e due, il visibile e l'invisibile. E io utilizzo la pittura per rendere visibile il pensiero (René François Ghislain Magritte, La firma in bianco), come mostra la Figura D.2.

Concludendo questa disamina sulla storia, la teoria e la pratica dei metodi per l'analisi tra la matematica e la filosofia, resta da rimarcare l'importanza di un cammino di riavvicinamento tra queste due discipline, come un esempio pregevole del superamento di un'infausta frattura, tra le scienze fisiche e naturali (in particolare, le scienze dure, a carattere matematico) e le scienze umane, a netto vantaggio di una crescita armoniosa della conoscenza (seguendo anche lo schema riportato nella Figura D.1) e per il conseguimento, reale e ben consolidato, del bene per l'umanità intera <sup>43</sup>.

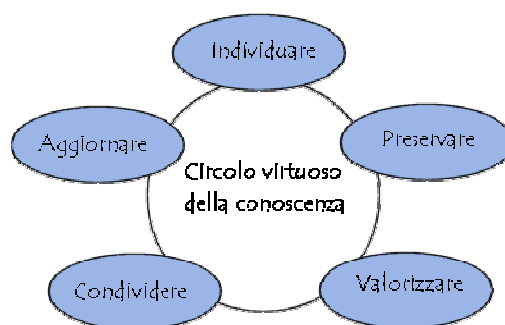


Fig. D.1 – Circolo virtuoso della conoscenza

<sup>42</sup> L'assunta esistenza di una grammatica profonda, base di ogni linguaggio naturale, è in accordo con le tesi generative di Avram Noam Chomsky (linguista ed attivista politico americano).

<sup>43</sup> In Italia, questa frattura è purtroppo particolarmente rilevante, da quanto, un secolo fa, a Bologna, si consuma un confronto culturale drammatico. Infatti il 6 aprile 1911, al Congresso della Società filosofica italiana, fondata e presieduta dal grande matematico Federigo Enriques, un aperto contrasto, con l'idealismo di Benedetto Croce e Giovanni Gentile, fa degenerare un confronto civile in una violenta disputa. Enriques che ritiene una filosofia degna di una società moderna, solo se pensata in stretta connessione con l'avanzare di tutte le scienze e della matematica, denuncia l'atteggiamento degli idealisti, nei confronti dei saperi scientifici, come antifilosofico. Ormai ad un secolo di distanza, ancora oggi, purtroppo è spesso molto difficile trovare, fattivi e fecondi, momenti comuni.



Fig. D.2 – René Magritte, Cavaliere dans le bois

## Ringraziamenti

Questo lavoro su Complementi di Statistica conclude una triade di lavori, iniziata con Elementi di Statistica e proseguita con Fondamenti di Statistica, dove con il primo autore di questo lavoro (e sempre con il secondo autore) si sono avvicinate rispettivamente Alberta Albertella e Vincenza Tornatore, tutte/i unite/i in doverosi ringraziamenti, rivolti a:

- ❑ gli ingegneri, allievi del corso di Geomatica e Trattamento delle Osservazioni del Tirocinio Formazione Attivo (TFA) per la classe A072 Topografia generale, nell'anno accademico 2014/2015: Biancamaria Borella, Carlo Chimento, Giambattista Cocci, Olga Iliceto, Emanuele Mazzola, Stefano Roberti, Scanderbergo Vaccaro che, durante il loro corso di studi, hanno contribuito ad editare il libro Fondamenti di Statistica di Giovanna Togliatti, come attività complementare per un seminario didattico, sul Trattamento delle osservazioni, integrativo al programma ufficiale per le Scuole medie superiori;
- ❑ l'architetto Ettore Scorsetti, già allievo del Corso di Dottorato di Ricerca in Geomatica e Infrastrutture del Politecnico di Milano che, per la stesura della propria Tesi di Dottorato sull'Analisi statistica del Disco di Festo, ha predisposto una valida sintesi sul tema del Controllo statistico della qualità, fonte intermedia della Appendice A di questo lavoro.

## BIBLIOGRAFIA ESSENZIALE <sup>44</sup>

Cunietti M. (1977): Le misure e il loro trattamento. CLUP, Milano.

De Finetti B. (1970): Teoria della probabilità – Sintesi introduttiva con appendice critica (vol. I e II). Giulio Einaudi Editore, Torino.

Frosini B.V. (1990): Introduzione alla statistica. La Nuova Italia Scientifica, Roma.

Montgomery D.C. (2000): Controllo statistico della qualità. McGraw-Hill, Milano.

Mood A.M., Graybill F.A., Boes D.C. (1988): Introduzione alla statistica. McGraw-Hill, Milano.

Ricci F. (1975): Statistica ed elaborazione statistica delle informazioni. Zanichelli, Bologna.

Sansò F. (1989): Il trattamento statistico dei dati. CLUP, Milano

Togliatti G. (1976): Fondamenti di statistica. Hoepli, Milano.

<sup>44</sup> Questa triade di lavori è debitrice nei confronti di alcuni articoli, del secondo autore (di questo lavoro) e di alcuni suoi amici e colleghi, di cui si dà elenco e ragione, nel prosieguo:

- ❑ per le note dei due lavori Elementi di Statistica Fondamenti di Statistica: Modelli ottimali per l'analisi dei dati e Validazione dei dati e dei modelli (nel sito: <http://intranet.dica.polimi.it/people/mussio-luigi/>);
- ❑ per l'Appendice B di questo lavoro ancora: Modelli ottimali per l'analisi dei dati (<http://intranet.dica.polimi.it/people/mussio-luigi/>) e Metodi statistici per la compensazione di reti nelle discipline geodetiche di Gianfranco Forlani (Ricerche di Geodesia Topografia e Fotogrammetria, volume 8, CLUP, Milano, 1990);
- ❑ per le Appendice C e D di questo lavoro: Ibridare la Geomatica di Alice Pozzoli, oltre al secondo autore di questo lavoro (in: Memorial "Giovanna Togliatti", a cura di Luigi Mussio, Federica Migliaccio, Franco Guzzetti e Livio Pinto, Politecnico di Milano / ASITA, Lecco, 2015).