

MODELLI OTTIMALI PER L'ANALISI DEI DATI

Luigi Mussio
Politecnico di Milano – DICA
Piazza L. da Vinci, 32 – 20133 Milano
Tel. 02–2399–6501, Fax. 02–2399–6602
e-mail luigi.mussio@polimi.it

RIASSUNTO

I principali aspetti algebrici e geometrici del metodo dei minimi quadrati seguono una breve introduzione sulla statistica descrittiva e la teoria della stima. I modelli lineari delle equazioni di condizione pure, con una dettagliata critica sul loro impiego, e delle equazioni di condizione con parametri, nello schema principe delle equazioni d'osservazione, precedono alcune generalizzazioni dei modelli lineari stessi. Fra esse, sono d'interesse le equazioni di condizione con parametri vincolati e le loro trasformazioni in equazioni di pseudo-osservazione sovrappesate, nonché l'impiego di equazioni di pseudo-osservazione qualsiasi, anche in importanti casi particolari complementari. Altri modelli generalizzati prendono in considerazione modelli più raffinati. Modelli non – lineari riferiti al modello funzionale, al modello stocastico ed alla norma, inoltre i problemi del condizionamento, dell'affidabilità e della robustezza, concludono la presentazione del metodo. Due appendici sviluppano, rispettivamente, la teoria della robustezza, i modelli partizionati e sequenziali. Un'altra appendice prende in considerazione modelli stocastici più raffinati e sviluppa il metodo della collocazione con e senza parametri.

PREMESSA

L'analisi dei dati a referenza spaziale, tempo varianti e non, quali, ad esempio, l'insieme dei problemi geodetici e geomatici (si veda, a riguardo, la loro illustrazione nei paragrafi a seguire), richiede, fra l'altro, il trattamento matematico, statistico e numerico delle osservazioni ed, in questo ambito, il trattamento statistico delle osservazioni fa uso, abbondante ed intensivo, del metodo dei minimi quadrati, delle sue generalizzazioni e delle estensioni a tecniche complementari, sia per quanto riguarda la stima dei parametri di modelli, che per quanto riguarda l'inferenza statistica. Il primo argomento è il tema centrale del presente lavoro, mentre il secondo, generalizzabile nell'inferenza statistica della normalità, giustapponibile all'inferenza statistica non-parametrica ed inseribile nell'ambito più ampio dell'analisi (statistica) multivariata, è considerato estraneo agli scopi del presente lavoro. La suddivisione e la scelta effettuate limitano, pertanto, la trattazione a seguire allo studio dei modelli ottimali, della stima dei parametri di questi e della loro comparazione e critica, in base alla capacità delle varie tecniche corrispondenti di rispondere ai bisogni dell'analisi dei dati a referenza spaziale. A questo punto, prima di entrare nel tema specifico, si presentano per completezza un riassunto della statistica descrittiva e brevi cenni di teoria della stima.

1. RICHIAMI DI STATISTICA

Un discorso preciso sulla statistica descrittiva si avvia con la definizione di variabile statistica e di variabile casuale, la postulazione di un'identità formale fra le stesse e la presentazione delle loro principali statistiche. Le variabili statistiche sono il risultato di esperimenti e, pertanto, sono concrete (ovvero costituite da dati reali od osservazioni, come la totalità dei dati a referenza spaziale, tempo varianti e non, quali, ad esempio, le misure geodetiche e geomatiche), finite (perché qualsiasi esperimento incontra evidenti limiti di spazio,

tempo ed altre condizioni limitative) e discrete (perché qualsiasi esperimento è eseguito con una determinata accuratezza). Conseguentemente esse sono caratterizzate da un insieme di valori argomentali (eventualmente raggruppati in classi), associati a frequenze elementari (assolute, come risultato di un conteggio, oppure relative, se la totalità è normalizzata ad uno) ed alle frequenze cumulate delle frequenze elementari.

Le variabili casuali sono modelli interpretativi e, pertanto, sono astratte (ovvero costituite da dati ideali od osservabili) ed, in generale, illimitate e continue (anche se, raramente, ad eccezione della teoria dei giochi, esse possono essere finite e discrete). Conseguentemente esse sono caratterizzate da un campo d'esistenza, associato ad una funzione densità di probabilità ed ad una funzione distribuzione di probabilità (comunemente detta: probabilità). L'identità formale fra variabili statistiche e variabili casuali discende dalla loro completa indistinguibilità, a valle della loro definizione. Allora la presentazione delle principali statistiche può essere eseguita congiuntamente per entrambe.

Per le variabili ad una dimensione, le principali statistiche rispondono alla quantizzazione delle idee di: centro, dispersione, simmetria e curtosi (comportamento delle code). Come noto, il centro può essere indicato tramite la moda, la mediana, le medie (aritmetica, geometrica, armonica, ponderata, potata, ecc.) od altro, la dispersione può essere valutata in base all'ampiezza, ai quantili, alla varianza, agli scarti assoluti medio o mediano, ecc., mentre gli indici di asimmetria e curtosi hanno, solitamente, poche varianti.

Per le variabili a due dimensioni, le principali statistiche (oltre a quelle monodimensionali marginali o condizionate) rispondono alla quantizzazione dell'idea di dipendenza. Come noto, dipendenza è un concetto molto generale che, fra totale e completa indipendenza e perfetta dipendenza (o dipendenza in legge), si articola in connessione (dipendenza vaga e generica), regressione (quasi-dipendenza funzionale) e correlazione (quasi-dipendenza lineare). Ancora numerosi sono gli indici ed i coefficienti che esprimono il grado della dipendenza o meno (si noti, a riguardo, come tutti siano normalizzati ad uno, assumendo anche valori negativi, fino a meno uno, se non intrinsecamente positivi).

Per le variabili a più di due dimensioni, a rigore, occorre continuare lo studio del loro raggrupparsi (come con gli indici di nuvolosità, ecc.). Tuttavia nel caso frequente in cui il modello interpretativo è fornito dalla variabile casuale normale, questo studio è del tutto superfluo. Si ricordi, inoltre, che detta variabile casuale è completamente caratterizzata dal vettore delle medie e dalla matrice di varianza-covarianza, cosa che rende superflue altre statistiche del centro, della dispersione e della dipendenza (covarianza comporta correlazione, ovvero dipendenza lineare e niente altro) e del tutto inutili le statistiche superiori (la variabile casuale normale è simmetrica e l'indice di curtosi vale, in ogni caso, tre).

Ulteriori vantaggi dell'adozione, quale modello interpretativo, della variabile casuale normale sono dati dall'invarianza della distribuzione di probabilità di detta variabile casuale, rispetto a trasformazioni lineari della variabile casuale stessa, e dell'ottimalità della stima dei parametri di modelli, supportati dalla variabile casuale normale, se le ipotesi di corrispondenza fra dati e modelli sono perfettamente soddisfatte (ovvero se i dati non sono affetti, in alcun modo, da dati anomali). A tutto ciò, si aggiunge la linearità dei sistemi da risolvere per la stima dei parametri di modelli, fatto di primaria importanza, in quanto solo i sistemi lineari ammettono, senza eccezioni e purché non-singolari, soluzioni esattamente determinabili, indipendentemente dal numero di equazioni ed incognite di cui si compongono.

La statistica descrittiva termina con alcuni teoremi limite. Fra questi il teorema di Bernoulli (o legge dei grandi numeri) mostra la convergenza, in probabilità, delle frequenze di una variabile statistica alle probabilità di

una corrispondente variabile casuale, mentre il teorema di Gauss (o limite centrale della statistica) mostra la convergenza, in legge, della combinazione lineare (ovvero delle somme, come caso particolare) di variabili casuali qualsiasi, purché aventi ciascuna dispersione comparabile con le altre, alla variabile casuale normale. I due teoremi giustificano, rispettivamente, la comparazione fra variabili statistiche e variabili casuali, al di là della sopracitata identità formale, e le operazioni di media aritmetica, ponderata o potata fra osservazioni dirette per aumentare la normalità del comportamento dei dati.

Brevi cenni di teoria della stima concludono quest'introduzione. La teoria della stima ha lo scopo di definire proprietà, caratteristiche e modalità della stima dei parametri di modelli, dove questi modelli, da interpretarsi, come già detto, quali variabili casuali, sono concepiti come popolazioni di dati ideali (od osservabili), costituenti un universo da cui estrarre dati reali (od osservazioni), costituenti, a loro volta, campioni da interpretare, pertanto, come le suddette variabili statistiche. L'operazione d'estrazione è detta campionamento e, se i dati sono fra loro incorrelati, come avviene, auspicabilmente, nelle osservazioni dirette di fenomeni, il campionamento è detto bernoulliano; schemi di campionamento più complessi, attinenti alle problematiche della progettazione, simulazione ed ottimizzazione, sono considerati estranei agli scopi del presente lavoro.

Le principali proprietà delle stime sono la correttezza, la consistenza, l'efficienza e la sufficienza che, in base al significato letterale dei nomi, significano capacità di stimare parametri il cui centro coincide con il centro dei parametri dell'intera popolazione, capacità di stimare parametri con precisione ed accuratezza sempre maggiori, ed al limite infinite, al crescere della numerosità del campione, capacità di stimare parametri qualitativamente migliori delle informazioni presenti direttamente nelle osservazioni, capacità di stimare parametri conservando tutta la ricchezza di informazioni già presente nelle osservazioni dirette. Una proprietà aggiuntiva delle stime, estranea al corpus della statistica tradizionale, è la robustezza, intesa come capacità di stimare parametri indistorti, nonostante la presenza di eventuali dati anomali.

Le caratteristiche delle stime forniscono spesso, contemporaneamente, le modalità operative per effettuare le stime stesse. Infatti sono operativi tanto il metodo della minima varianza, ottimale per la statistica tradizionale, quanto diverse procedure robuste, certamente subottimali, ma capaci di evitare indesiderate distorsioni nelle stime. Altrettanto operativi sono il metodo della massima verosimiglianza ed il metodo dei minimi quadrati, una particolarizzazione del primo nel caso in cui le stime avvengono in ambito lineare ed il modello interpretativo è fornito dalla variabile casuale normale. Si ricordi che, in questo caso, si ha l'ottimalità delle stime, in quanto tanto il metodo dei minimi quadrati, quanto quello della massima verosimiglianza, da cui discende, danno risultati perfettamente coincidenti con il metodo della minima varianza.

Tutto ciò conferma l'adozione del metodo dei minimi quadrati per il trattamento statistico delle osservazioni, agevola le sue generalizzazioni ed estensioni a tecniche complementari e giustifica un modo di procedere che prevede la centralità di detto metodo e riconduce ad esso, per quanto possibile, importanti tecniche complementari (cluster analysis, regressione multipla, analisi di varianza, delle componenti di varianza e della struttura di covarianza, analisi fattoriale, procedure robuste). Si noti, in quest'ambito, il ruolo fondamentale ed indispensabile giocato dall'indissolubilità del legame fra un certo tipo di statistiche classiche elementari, la normalità e la linearità, per quanto riguarda tanto la definizione statistica delle metodologie, quanto la loro applicazione con elevate capacità risolutive in appropriati algoritmi numerici ¹.

¹ Il capitolo successivo della statistica riguarda l'inferenza statistica e tratta del controllo e confronto d'ipotesi. Essa si articola in una lunga serie di test statistici, semplici e multipli, della normalità e non – parametrici (ovvero distribution – free). Tutti i test sono fondati sulla selezione di opportune distribuzioni di probabilità per spiegare il comportamento delle diverse statistiche campionarie che, avendo

2. RICHIAMI DI ALGEBRA.

Il metodo dei minimi quadrati ha, certamente, il suo fondamento nell'ambito della statistica ed, in particolare, come già detto, della teoria della stima; tuttavia alcune sue caratteristiche formali e interpretative fanno riferimento, rispettivamente, a strutture algebriche ed a rappresentazioni geometriche. La geometria dei minimi quadrati è costituita dallo spazio delle osservabili e, se questi sono definiti, dallo spazio dei parametri. Le osservazioni nello spazio delle osservabili sono punti in detto spazio. D'altra parte poiché, in generale, le osservabili sono legate fra loro da relazioni di condizione, lo spazio delle osservabili compatibili fra loro è un sottospazio del suddetto spazio delle osservabili.

Questo sottospazio è, in generale, non-lineare ed è lineare, solo se lineari sono le relazioni di condizione fra le osservabili. Nel caso in cui il sottospazio delle osservabili compatibili fra loro è parametrizzato, l'insieme dei parametri costituisce uno spazio in corrispondenza biunivoca con il suddetto sottospazio. Più in generale, la corrispondenza biunivoca si ha fra un sottospazio dello spazio dei parametri ed il sottospazio delle osservabili compatibili fra loro, essendo il sottospazio complementare dei parametri in corrispondenza biunivoca con lo spazio dei vincoli (o delle pseudo-osservazioni) basati su informazioni a priori.

Con specifico riferimento all'insieme dei problemi geodetici e geomatici, i due casi sopraesposti individuano due distinte classi di problemi, noti nella letteratura specialistica, rispettivamente, come problemi d'interpolazione/approssimazione e di tipo reticolare. I primi fanno riferimento alla ricostruzione di linee, superfici ed oggetti, alla centratura (matching) di figure ed oggetti, alla descrizione di contorni di figure ed oggetti. I secondi, invece, fanno riferimento alla compensazione di reti di tipo geodetico, alle loro varie scale, natura ed estensione, nonché ai problemi fotogrammetrici dell'orientamento di immagini, telerilevate e non, e del posizionamento di punti. A complemento di quanto sopra esposto, si ricorda che è un problema della prima classe l'interpolazione di serie temporali, mentre sono problemi della seconda classe la compensazione di reti di trasporto, qualunque sia la loro struttura e destinazione, entrambi d'interesse per l'analisi di dati a referenza spaziale, tempo varianti e non.

La determinazione di un valore delle osservabili, compatibile con le relazioni di condizione ed in corrispondenza al valore delle osservazioni, avviene per proiezione nel sottospazio delle osservabili compatibili fra loro, a partire dal suddetto valore delle osservazioni.

La direzione della proiezione è sempre ortogonale, in generale dopo aver equiponderato le osservabili e, se dal caso, averle sbiancate (cioè rese indipendenti fra loro); ciò significa che lo spazio ove effettivamente avviene la proiezione ortogonale è uno spazio ortonormalizzato affine allo spazio delle osservabili (infatti la trasformazione affine ha cambiato, solamente, la scala e la pendicolarità delle direzioni coordinate). Il punto piede della proiezione ortogonale sul sottospazio delle osservabili compatibili fra loro esiste ed è unico se questo sottospazio è lineare, mentre non si ha alcuna garanzia né sull'esistenza, né sull'unicità, se questo sottospazio è non-lineare.

Ciò da ragione, fra l'altro alla scelta di operare iterativamente, in ambito locale, su una porzione lineare, ottenuta per linearizzazione del suddetto sottospazio. A sua volta, la linearizzazione è possibile, geometricamente, se si ha corrispondenza biunivoca fra spazi non – lineari e porzioni linearizzate e, statisticamente, se la distribuzione di probabilità delle osservabili è, ivi, ben concentrata.

scelto e fissato un livello di significatività ed eventualmente una potenza del test, sono partizionate in una zona dove vale l'ipotesi fondamentale, in una o due zone critiche (dove possono valere alcune ipotesi alternative)e, se del caso, in piccole regioni del dubbio. Per una prima trattazione di questa parte, ad eccezione dei test multipli della normalità e non, come pure per quanto riguarda la statistica descrittiva, si rimanda all'ultima appendice di questo lavoro, oltreché alla vastissima letteratura specialistica, spesso scritta e disponibile anche in italiano.

L'algebra dei minimi quadrati fa riferimento all'algebra lineare delle matrici ed, in particolare, si fonda sul concetto di matrice inversa generalizzata, definita la quale è immediato definire, a partire dal valore delle osservazioni, tanto il valore delle stime delle osservabili compatibile con le relazioni di condizione, quanto il valore delle stime dei parametri, se questi sono definiti.

Brevi richiami sulle strutture algebriche sono riportati di seguito e, a queste, fanno riferimento importanti proprietà dell'algebra lineare delle matrici. Ulteriori proprietà, teoremi e corollari sulle strutture algebriche, in quanto propri dell'algebra (formale), sono considerati estranei agli scopi del presente lavoro.

Def. di gruppo. Si chiama gruppo (gruppo abeliano) un insieme munito di una legge di composizione, comunemente detta prodotto (somma), che soddisfa i seguenti postulati:

- ❑ l'elemento ottenuto, operando, mediante la legge di composizione, su due elementi qualsiasi del gruppo, appartiene allo stesso gruppo, ovvero il gruppo è chiuso rispetto al prodotto (alla somma);
- ❑ la legge di composizione gode della proprietà associativa (inoltre gode della proprietà commutativa, se il gruppo è abeliano);
- ❑ nel gruppo esiste un elemento neutro che funziona da unità (zero), rispetto al prodotto (alla somma);
- ❑ ogni elemento del gruppo ammette un elemento inverso (opposto), cioè un elemento, appartenente allo stesso gruppo, che, a prodotto (in somma) con l'elemento dato, dà l'unità (lo zero).

Def. di anello. Si chiama anello un insieme munito di due leggi di composizione, comunemente dette somma e prodotto, che soddisfa i seguenti postulati:

- ❑ l'insieme è un gruppo abeliano rispetto alla somma;
- ❑ l'insieme è un semigruppato rispetto al prodotto, perché ogni elemento dell'anello non ammette, in generale, un elemento inverso, cioè un elemento, appartenente allo stesso anello, che, a prodotto con l'elemento dato, dà l'unità;
- ❑ valgono le proprietà distributive (destra e sinistra) del prodotto rispetto alla somma.

Def. di corpo (campo). Un anello in cui gli elementi, diversi dallo zero, formano un gruppo (gruppo abeliano) rispetto all'operazione prodotto è chiamato corpo (campo); infatti ogni elemento del corpo (campo), tolto lo zero, ammette sempre un elemento inverso, cioè un elemento, appartenente allo stesso corpo (campo), che, a prodotto con l'elemento dato, dà l'unità.

Def. di modulo (unitario), spazio vettoriale, algebra (con divisione). Si chiamano, nell'ordine: modulo (unitario) sopra un anello e spazio vettoriale sopra un corpo o un campo; algebra (con divisione) sopra le medesime strutture algebriche,

- ❑ gli insiemi costituiti da: nel primo caso, un gruppo abeliano;
 nel secondo caso, un anello (o un corpo o un campo),

per cui, definito un prodotto esterno fra gli elementi dei due insiemi, siano soddisfatti i seguenti postulati:

- l'elemento ottenuto, operando, mediante il prodotto esterno, su due elementi qualsiasi dei due insiemi, appartiene al primo insieme, ovvero questa struttura algebrica è chiusa rispetto al prodotto esterno definito;
- valgono, per entrambi gli insiemi, le proprietà distributive del prodotto esterno rispetto alla somma interna.
- il prodotto esterno gode della proprietà associativa con il prodotto interno fra gli elementi del secondo insieme;
- il prodotto esterno gode della proprietà associativa con il prodotto interno fra gli elementi del primo insieme, se questo è un'algebra (con divisione);
- nel secondo insieme esiste un elemento neutro che funziona da unità rispetto al prodotto esterno, se il primo insieme è, quantomeno, un modulo unitario;

Si osservi a proposito che l'insieme delle matrici, di una data dimensione, costituisce:

- uno spazio vettoriale sopra il campo dei numeri reali, dove il prodotto esterno è il prodotto definito scalare per vettore;
- un modulo (unitario) sopra l'anello delle matrici quadrate, di dimensioni compatibili, dove il prodotto esterno è definito con il consueto prodotto matriciale righe per colonne;
- un insieme qualsiasi (detto: raggruppamento o gruppoide) sopra il gruppo abeliano delle matrici qualsiasi, di dimensioni compatibili, dove il prodotto esterno è lo stesso sopracitato, non essendo chiusa, come noto, questa operazione.

La non-chiusura del raggruppamento o gruppoide delle matrici qualsiasi, di dimensioni compatibili, rispetto al prodotto esterno definito con il consueto prodotto matriciale righe per colonne, fa sì che l'algebra lineare delle matrici non sia affatto un'algebra (con divisione), nemmeno considerando il modulo unitario delle matrici quadrate definite positive, di dimensioni compatibili. Infatti il sottoinsieme delle matrici quadrate definite positive, di dimensioni compatibili, a sua volta, non gode della proprietà della chiusura, rispetto all'operazione di somma. Allora accanto alla matrice regolare delle matrici quadrate definite positive, si definiscono più matrici inverse generalizzate per le matrici qualsiasi (comprese le matrici quadrate non definite positive), tutte coincidenti con l'unica matrice inversa regolare, quando questa esiste².

¹ Si chiama matrice definita positiva, una matrice quadrata e simmetrica Q tale che risulti :

$$\underline{u}^T Q \underline{u} < 0 \quad \forall \underline{u}$$

$$\underline{u}^T Q \underline{u} = 0 \quad \rightarrow \underline{u} = 0$$

Si chiama matrice inversa generalizzata di una qualsiasi matrice rettangolare $A(m,n)$, una matrice $A^-(n,m)$ per la quale vale la relazione a) delle seguenti relazioni:

$$A A^- A = A$$

$$A^- A A^- = A^-$$

$$(A^- A)^T = A A^-$$

Il calcolo numerico di matrici inverse generalizzate, quando queste non coincidono con le matrici originarie (come per le matrici proiettore), è spesso non agevole. Tuttavia la postulazione e verifica di particolari matrici inverse generalizzate, dette pseudo-inverse, ricavabili anche per altra via nell'ambito della statistica ed, in particolare, della teoria della stima, risponde positivamente alla bisogna.

Conseguentemente tanto nel caso equazioni di condizione pure (sole relazioni di condizione per le osservabili che definiscono il sottospazio non parametrizzato delle osservabili compatibili fra loro), quanto nel caso di equazioni di condizione con parametri, eventualmente vincolati o pseudo-osservati in parte o in toto, (quando il suddetto sottospazio è parametrizzato), una particolare matrice inversa generalizzata, detta pseudo-inversa, è definita.

Nel primo caso, essendo le equazioni di condizione in numero inferiore alle osservabili, la matrice originaria è una matrice rettangolare larga e la matrice inversa generalizzata sarà una matrice rettangolare lunga, di dimensioni compatibili:

$$B \rightarrow P^{-1} B^T (B P^{-1} B^T)^{-1}$$

Nel secondo caso, essendo le equazioni di osservazione, vincolo o pseudo-osservazione (caso solo apparentemente particolare che, invece, nulla toglie alla generalità del problema, come sarà mostrato nel prosieguo) in numero superiore ai parametri, la matrice originaria è una matrice rettangolare lunga e la matrice inversa generalizzata sarà una matrice rettangolare larga, di dimensioni compatibili:

$$A \rightarrow (A^T P A)^{-1} A^T P$$

La matrice P rappresenta, in entrambi i casi, usualmente la matrice dei pesi delle osservazioni e, più in

$$(A A^T)^T = A^T A$$

Se vengono soddisfatte le prime due condizioni, la matrice inversa generalizzata è detta simmetrica e riflessiva.

Se vengono soddisfatte tutte e quattro le suddette condizioni, la matrice si chiama matrice pseudo-inversa, nota anche come inversa di Moore-Penrose, generalmente, indicata mediante il simbolo A^+ . L'inversa di Moore-Penrose è unica e gode, inoltre, delle seguenti proprietà:

$$(A^+)^+ = A$$

$$(A^+)^T = (A^T)^+$$

$$R(A) = R(A^+)$$

$$A^+ = (A^T A)^+ A^T$$

$$A^+ = A^T (A A^T)^+$$

La matrice pseudo-inversa di una matrice A quadrata regolare ($\det A \neq 0$) coincide con l'inversa regolare di Cayley:

$$A^+ = A^{-1}$$

generale, la matrice dei cofattori, ottenuta per inversione della matrice di varianza–covarianza delle osservazioni stesse, se queste non sono indipendenti fra loro.

L'algebra lineare delle matrici affianca, poi, ai consueti prodotto scalare per vettore (o matrice) e prodotto matriciale righe per colonne, altri prodotti speciali. Fra questi, il prodotto di Kronecker costituisce la generalizzazione del primo nel prodotto fra matrici, mentre il prodotto di Hadamard è giustapponibile al secondo, operando con la stessa semplicità dell'operazione di somma, sempre nel prodotto fra matrici.

Si chiama prodotto di Kronecker fra due matrici A e B , aventi rispettivamente dimensioni m, n e k, l , la matrice: $C = A \otimes B$, avente dimensioni km, ln , ottenuta moltiplicando (con la regola del prodotto di una matrice per uno scalare) la matrice B per tutti gli elementi della matrice A :

$$C = A \otimes B = [a_{ij} B] \quad \forall i, j$$

Si chiama prodotto di Hadamard fra due matrici A e B , aventi le stesse dimensioni m, n , la matrice: $C = A * B$, avente ancora le stesse dimensioni, ottenuta moltiplicando (con la regola del prodotto dell'aritmetica) ogni elemento della matrice B per il corrispondente elemento della matrice A :

$$C = A * B = [a_{ij} b_{ij}] \quad \forall i, j$$

Altri prodotti speciali (ad es., il prodotto di Katri Rao) sono definiti, talvolta, per usi specifici: resta da citare la definizione dell'operatore *vec*, capace di disporre in un vettore, una dopo l'altra, tutte le colonne di una matrice.³

3. CLASSI DI PROBLEMI

I problemi ai minimi quadrati si presentano, in generale e nell'ambito specifico delle discipline geodetiche e geomatiche, usualmente ripartiti in due classi fondamentali:

- ❑ problemi reticolari (o di compensazioni di reti);
- ❑ problemi d'interpolazione ed approssimazione di campi di punti.

Le stesse due classi si incontrano anche in problemi affini, quali ad esempio:

- ❑ i campionamenti delle osservazioni, l'ottimizzazione della configurazione di rilevamento e/o dello schema di misura, oppure dei pesi delle osservazioni;
- ❑ la "cluster analysis", l'analisi di varianza, la regressione multipla, l'analisi fattoriale (o studio delle

³ Si chiama prodotto di Katri–Rao fra due matrici A e B , aventi rispettivamente dimensioni m, n e k, n la matrice: $C = A \odot B$, avente dimensioni km, n ottenuta moltiplicando (con la regola del prodotto di un vettore per uno scalare) ogni colonna della matrice B per tutti gli elementi della corrispondente colonna della matrice:

$$C = A \odot B = [\underline{a}_j \otimes \underline{b}_j] = [a_{ij} \underline{b}_j] \quad \forall i, j$$

componenti principali);

- lo studio dell'affidabilità delle osservazioni e le procedure di validazione dei dati e di stima dei parametri con procedure robuste.

In ogni caso, tutti i problemi minimi quadrati possono essere interpretati, topologicamente, come un grafo, dove:

- le osservazioni, i vincoli e le pseudo-osservazioni sovrappesate e non, come pure le informazioni a priori, le osservabili secondarie e le condizioni numeriche di regolarizzazione, costituiscono i lati del grafo;
- i parametri principali ed ausiliari (o di servizio) costituiscono i nodi dello stesso grafo.

Si noti, a riguardo, come l'interpretazione data della topologia sia indispensabile per una corretta comprensione dei casi e sottocasi in cui si articolano le suddette classi fondamentali.

Le due classi fondamentali, già precedentemente enunciate, si articolano in vari e svariati casi e sottocasi, illustrati, dettagliatamente, nel prosieguo.

I problemi reticolari (o di compensazione di reti) presentano come osservabili:

- differenze prime dei parametri;
- funzioni delle differenze prime dei parametri.

Il primo caso ha numerosi esempi, anche fuori dalle discipline geodetiche e geomatiche:

- discretizzazione di equazioni differenziali del primo ordine, tipiche della fisica, della chimica e delle scienze della terra;
- problemi di trasporto: schemi di circuitazione, traffico, circolazione e transazione, reti di comunicazione, distribuzione e telecomunicazione ed è costituito, per le suddette discipline, dalle reti di differenza di potenziale.

Il secondo caso è, invece, tipico delle discipline geodetiche e geomatiche, anche se non esclusivo (a questo caso, infatti, fanno riferimento ben particolari discretizzazioni di equazioni differenziali, sempre riferite ai sopracitati raggruppamenti di discipline fisiche e naturalistiche), e si articola nei seguenti sottocasi:

- l'informazione fluisce completa, bidirezionalmente, come nelle reti di differenza di potenziale, lungo ogni lato del grafo:

$$\left\{ (P_i, P_j) = Q_k, \quad k = 1, m \right\}$$
$$\left\{ P_i, \quad i = 1, n < m \right\}$$

- l'informazione fluisce completa, unidirezionalmente, lungo ogni lato del grafo, costituendo nel suo fluire almeno un albero sul grafo stesso;

- l'informazione è irradiata, in modo completo, da alcuni nodi verso altri (senza ritorno), senza che né i primi, né i secondi si scambino alcuna informazione, costituendo nel suo fluire tanti alberi (costituiti da un solo livello, oltre la radice) sul grafo stesso, quanti sono i nodi d'emanazione:

$$\left\{ (P_i, Q_j) = R_k, \quad k = 1, m \right\}$$

$$\left\{ Q_j, \quad j = 1, n < m \right\}$$

$$\left\{ P_i, \quad i = 1, l < n \right\}$$

- l'informazione è irradiata, in modo parziale, nelle stesse condizioni del sottocaso precedente, cosa che richiede l'individuazione di due o più co-alberi (sempre costituiti da un solo livello, oltre le radici) capaci di completare l'informazione trasmessa;
- l'informazione è irradiata, in modo parziale, senza restituzioni particolari.

Le osservabili differenze seconde dei parametri e loro funzioni richiedono la complessa sostituzione dei lati del grafo con triangoli fra i tre nodi interessati. Le osservabili differenze di ordine superiore e loro funzioni fanno riferimento, addirittura, a poligoni fra tutti i nodi coinvolti, cosa che rende la loro analisi ancora più complessa. Per queste ragioni, ad eccezione della discretizzazione di equazioni differenziali di secondo ordine o di ordine superiore e di loro trasformazioni funzionali, la loro adozione è estremamente rara.

Per quanto riguarda la determinazione del numero di parametri principali, nel caso in cui i problemi ai minimi quadrati adottino lo schema principe delle equazioni d'osservazione, questo è sempre tale da determinare difetti di rango e singolarità del sistema da risolvere per cui sono necessari vincoli o pseudo-osservazioni sovrappesate.

I problemi d'interpolazione ed approssimazione di campi di punti presentano come osservabili funzioni dirette dei parametri principali, il cui numero, sempre nel caso in cui si voglia adottare il suddetto schema principe delle equazioni d'osservazione, non è mai tale da determinare difetti di rango e singolarità del sistema da risolvere. A tutto ciò, fanno eccezione eventuali problemi di sovra-parametrizzazione, rispetto al campionamento delle osservazioni effettuate, per cui sono indicate condizioni numeriche di regolarizzazione. Esempi di problemi di interpolazione ed approssimazione di campi di punti sono dati da:

- ricostruzione (fitting) di linee, superfici, ipersuperfici aventi come dominio lo spazio 3D:

$$s = f(x)$$

$$s = f(x, y)$$

$$s = f(x, y, z)$$

- descrittori di forma (form descriptors): contorni di figure (piane e/o gobbe), superfici (chiuse) di oggetti:

$$x = f(s) \quad ; \quad y = g(s) \quad ; \quad [z = h(s)]$$

$$x = f(s, t) \quad ; \quad y = g(s, t) \quad ; \quad z = h(s, t)$$

- centratura (matching) di segmenti, figure (immagini, mappe, disegni), oggetti (compresi modelli virtuali 3D) comunque conformati:

$$\begin{aligned}
 x &= f(u) \\
 x &= f(u, v) \quad ; \quad y = g(u, v) \\
 x &= f(u, v, w) \quad ; \quad y = g(u, v, w) \quad ; \quad z = h(u, v, w)
 \end{aligned}$$

I problemi d'interpolazione ed approssimazione di campi di punti, relativi alla ricostruzione di linee sono, ovviamente, assimilabili a quelli dello studio delle serie temporali storiche o di breve periodo, oppure frutto di simulazioni. Inoltre lo studio di serie temporali congiunto alla soluzione dei problemi reticolari (o di compensazione di reti) e/o d'interpolazione ed approssimazione di campi di punti, illustrati in precedenza, permette indagini accurate sugli aspetti dinamici delle osservabili a referenza spaziale di cui ai suddetti problemi, dando un'interpretazione unitaria a dati spazio-varianti, tempo-varianti.

Limitatamente alle discipline geodetiche e geomatiche, mentre le equazioni d'osservazione dei problemi reticolari (e di compensazioni di reti) fanno uso, in generale, di modelli grigi dedotti dalla geometria del problema in esame, le equazioni d'osservazione dei problemi d'interpolazione ed approssimazione di campi di punti fanno uso, in generale, di modelli neri.

Come noto, una vasta gamma di metodi deterministici e/o stocastici risponde positivamente alla bisogna. I primi annoverano fra i più comunemente impiegati:

- l'interpolazione polinomiale;
- il metodo degli elementi finiti e l'interpolazione con funzioni splines;
- l'analisi di Fourier, nel dominio delle frequenze;
- lo studio, sempre nel dominio delle frequenze, con ondine (wavelets).

I secondi prevedono l'interpretazione dei fenomeni in istudio come realizzazioni di un processo stocastico:

- stime di covarianza, filtraggio cross-validazione e predizione;
- studio della geometria frattale.

Uno studio dettagliato di esempi particolari e significativi di problemi reticolari (o di compensazioni di reti) può essere effettuato, nell'ambito delle discipline geodetiche e geomatiche, solo facendo riferimento a discipline specifiche, quali la geodesia, la navigazione, la topografia, la fotogrammetria ed il telerilevamento.

Al contrario, uno studio dettagliato di esempi particolari e significativi di problemi d'interpolazione ed approssimazione di campi di punti richiede anche uno studio dei modelli neri.

Per una migliore comprensione si tenga presente che l'insieme delle quantità osservate è sempre costituito da quattro parti distinte:

- le informazioni topologiche, ovvero i lati del grafo che indicano le connessioni esistenti fra i nodi del grafo stesso;
- le informazioni geometriche, ovvero la posizione ed altre caratteristiche degli stessi nodi;

- le informazioni metrologiche, ovvero le osservazioni (o quantità osservate) realmente effettuate;
- le informazioni stocastiche, ovvero la precisione delle osservazioni e le eventuali correlazioni fra queste.

Infatti questo insieme, altrimenti detto: base di dati provenienti da operazioni di misura, con riferimento a ciascuna delle sopraccitate quattro parti distinte, produce nei problemi ai minimi quadrati (come pure negli altri sopraccitati problemi affini), rispettivamente:

- la matrice disegno simbolica;
- la matrice disegno numerica;
- il vettore termine noto delle equazioni d'osservazione;
- la matrice di varianza–covarianza (a priori) delle quantità osservate o, più comunemente, se non esistono correlazioni fra le stesse quantità osservate, la matrice dei pesi.

Si noti che, con la sola eccezione delle osservazioni realmente effettuate, tutto quanto può essere noto già prima di compiere una sola osservazione. Da ciò derivano tutti i problemi di ottimizzazione della matrice di varianza–covarianza dei parametri:

- intervenendo nella matrice disegno per decidere sull'effettuazione o meno di ciascuna osservazione (1° ordine);
- sulla matrice di varianza–covarianza (a priori) delle quantità osservate per stabilire, note le osservazioni da effettuarsi, le precisioni delle stesse (2° ordine);
- su opportune parziali combinazioni dei due casi precedenti (3° ordine),

avendo cura di controllare, in ogni caso, l'affidabilità delle osservazioni, quale garanzia, sufficiente minimale, che il lavoro intrapreso, qualsiasi esso sia, risulti svolto a regola d'arte.

4. EQUAZIONI DI CONDIZIONE PURE

Un problema minimi quadrati, costituito da relazioni di condizione fra osservabili presenta il seguente modello funzionale lineare o linearizzato:

$$By + \Delta = 0$$

dove B rappresenta la matrice dei coefficienti e Δ il vettore dei termini noti delle equazioni di condizione pure (altrimenti detto, con un'espressione più antica, chiusura), ed il seguente modello stocastico:

$$E(y) = \bar{y}$$

$$D(y) = C_{yy} = \sigma_0^2 Q_{yy}$$

dove la matrice di varianza–covarianza delle osservazioni è, usualmente, nota a meno di un fattore di proporzionalità σ_0^2 (modello di Gauss–Markov: $Q_{yy} = Q_{y_0 y_0}$).

D'altra parte, come ogni problema concreto, esso contiene un campione di m osservazioni y_0 che si suppongono estratte dalle suddette osservabili, in generale, non soddisfacenti le relazioni di condizione fra le osservabili stesse. Allora opportune variabili correzione o scarti-residui \hat{v} collegano le osservazioni alle stime delle osservabili compatibili con le relazioni di condizione \hat{y} . La fondamentale indeterminazione del problema posto: esistono, cioè, infinite correzioni possibili, richiede l'introduzione di un criterio di scelta, a sua volta largamente arbitrario, che definisce, così, il metodo dei minimi quadrati. La sua formalizzazione matematica fa sì che una norma quadratica degli scarti-residui sia minimizzata, tenendo conto tanto del modello funzionale (tramite un vettore ausiliario di moltiplicatori di Lagrange), quanto del modello stocastico, già definiti in precedenza:

$$\begin{aligned}\phi &= \frac{1}{2}(\hat{y} - y_0)^T Q_{y_0 y_0}^{-1} (\hat{y} - y_0) + \lambda^T (B\hat{y} + \Delta) = \\ &= \frac{1}{2}\hat{v}^T P\hat{v} + \lambda^T (B\hat{v} + By_0 + \Delta) = \min\end{aligned}$$

La sostituzione della matrice dei pesi delle osservazioni P alla matrice dei cofattori, ottenuta per inversione della matrice di varianza-covarianza delle osservazioni stesse, ha carattere pratico e nulla toglie alla generalità del problema (in nessun caso, infatti, si farà uso dell' essere, solitamente, la matrice dei pesi delle osservazioni una matrice diagonale).

Si riportano di seguito il valore delle stime delle osservabili compatibili con le relazioni di condizione e delle correzioni o scarti-residui, facendo uso della nozione di matrice inversa generalizzata (definita nel paragrafo precedente, relativamente al primo dei due casi esposti) ed omettendo inutili passaggi algebrici elementari:

$$\begin{aligned}\hat{v} &= -P^{-1}B^T(BP^{-1}B^T)^{-1}(By_0 + \Delta) = \hat{y} - y_0 \\ \hat{y} &= y_0 + \hat{v} = y_0 - P^{-1}B^T(BP^{-1}B^T)^{-1}(By_0 + \Delta)\end{aligned}$$

La nota legge di propagazione della varianza-covarianza permette di ricavare, facilmente, le matrici di varianza-covarianza delle stime effettuate:

$$\begin{aligned}C_{\hat{v}\hat{v}} &= \hat{\sigma}_0^2 P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1} = C_{y_0 y_0} - C_{\hat{y}\hat{y}} \\ C_{\hat{y}\hat{y}} &= C_{y_0 y_0} - C_{\hat{v}\hat{v}} = \hat{\sigma}_0^2 \left(P^{-1} - P^{-1}B^T(BP^{-1}B^T)^{-1}BP^{-1} \right) \\ C_{\hat{y}\hat{v}} &\equiv 0\end{aligned}$$

dove la stima del fattore di proporzionalità (sigma zero):

$$\hat{\sigma}_0^2 = \frac{\hat{v}^T P \hat{v}}{\nu}$$

è ottenuta imponendo, direttamente, la sua correttezza:

$$E(v^T P v) = \text{Tr} P E(v v^T) = \sigma_0^2 \text{Tr} P Q_{vv} = \sigma_0^2 v$$

Tutte le stime godono delle proprietà di correttezza e consistenza; si ha, inoltre, efficienza delle stime rispetto alle osservazioni e, ovviamente, sufficienza delle stesse in termini di contenuto statistico (manca, invece, la robustezza, proprietà antitetica al metodo dei minimi quadrati).

All'elegante formulazione ed apparente semplicità del metodo dei minimi quadrati con equazioni di condizione pure si contrappongono notevoli difficoltà di ordine pratico: la determinazione delle equazioni di condizione pure non è, quasi mai, automatica ed, a volte, complessa.

A mo' d'esempio, le equazioni di condizione pure tra tutti gli angoli di un triangolo, scomposto in tre triangoli (avendo individuato un punto al suo interno e collegato quest'ultimo con i suoi vertici) sono, oltre alle chiusure dei triangoli ed al giro d'orizzonte, anche la congruenza della figura scomposta. Infatti muovendo i singoli lati parallelamente a se stessi, gli angoli non cambiano perché corrispondenti. Le figure 4.1 e 4.2 mostrano l'esempio proposto e la sua degenerazione; da qui applicando, iterativamente, il teorema dei seni, si giunge alla determinazione dell'ultima equazione di condizione pura:

$$\sin \alpha_1 \sin \beta_1 \sin \gamma_1 = \sin \alpha_2 \sin \beta_2 \sin \gamma_2$$

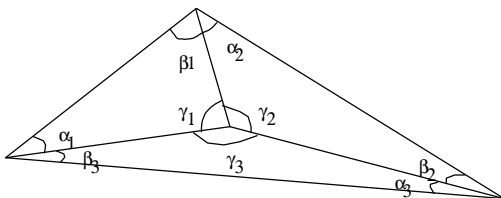


Fig. 4.1

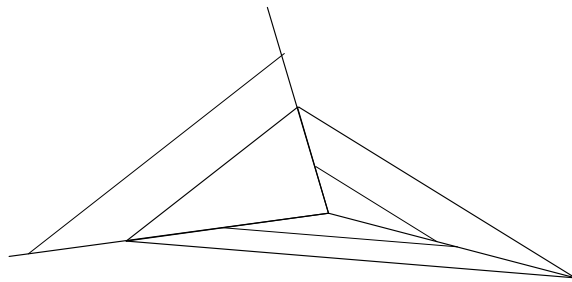


Fig. 4.2

Si badi invece che la chiusura del triangolo esterno è una combinazione lineare di tutte le altre chiusure; pertanto la corrispondente equazione di condizione pura è, del tutto, superflua.

Inoltre, anche le equazioni di condizione pura tra tutti gli angoli di un quadrilatero sono, oltre alle chiusure del quadrilatero e di due triangoli non provenienti dalla stessa scomposizione del quadrilatero stesso, anche la congruenza della figura scomposta. Infatti muovendo ancora i singoli lati parallelamente a se stessi, gli angoli non cambiano, perché corrispondenti. Le figure 4.3 e 4.4 mostrano l'esempio proposto e la sua degenerazione; da qui applicando iterativamente il teorema dei seni, si giunge alla determinazione dell'ultima equazione di condizione pura:

$$\sin \alpha \sin \gamma \sin \varepsilon \sin \vartheta = \sin \beta \sin \delta \sin \eta \sin \zeta$$

Si badi invece che ancora le chiusure degli altri due triangoli sono una combinazione lineare delle altre chiusure; pertanto le corrispondenti equazioni di condizione pure sono, del tutto, superflue.

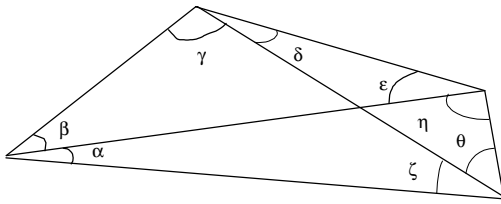


Fig. 4.3

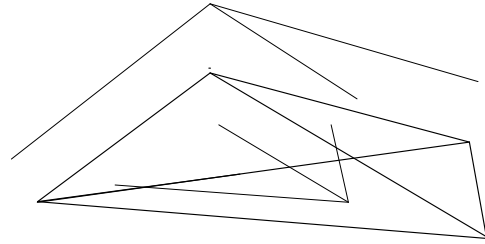


Fig. 4.4

Analogamente, l'equazione di condizione pura tra tutti i lati di un quadrilatero qualsiasi con un punto interno (si veda, a riguardo, la figura 4.5), si ricava dalla chiusura a 2π di tutti gli angoli del quadrilatero (o indifferentemente di tutti gli angoli del giro d'orizzonte nel punto interno), dove la determinazione di un qualsiasi angolo ϑ di un triangolo, dati i suoi lati (t opposto al vertice, inoltre r ed s), può essere ottenuta risolvendo, rispetto al coseno dell'angolo, il teorema di Carnot, ovvero facendo uso delle formule di Briggs:

$$\cos \vartheta = \frac{r^2 + s^2 - t^2}{2rs}$$

$$\sin \frac{\vartheta}{2} = \sqrt{\frac{(p-r)(p-s)}{rs}}$$

$$\cos \frac{\vartheta}{2} = \sqrt{\frac{p(p-t)}{rs}}$$

essendo p il semiperimetro del triangolo

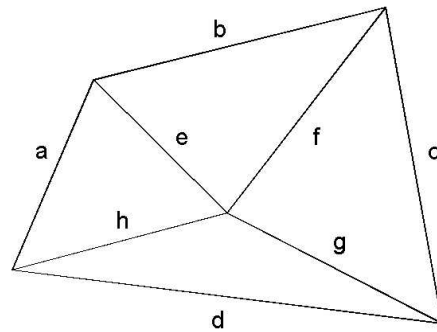


Fig. 4.5

Infine, l'equazione di condizione pura tra i lati e le diagonali di un quadrilatero qualsiasi (oppure tra tutti i lati di un triangolo con un punto interno) si ricava dapprima applicando ai quattro triangoli, in cui il quadrilatero è scomposto dalle due diagonali (oppure il triangolo è scomposto dai lati interni), la formula di Erone (per determinare l'area di un triangolo dati i suoi lati) e, successivamente, addizionando, a due a due, le aree ottenute così da ottenere due volte l'area del quadrilatero (oppure sottraendo dall'area del triangolo maggiore una delle aree di un triangolo interno per ottenere le aree degli altri due); ovviamente, per l'univocità delle figure, le due aree devono essere uguali fra loro:

$$\sqrt{p(p-a)(p-b)(p-e)} \pm \sqrt{q(q-c)(q-d)(q-e)} =$$

$$= \sqrt{r(r-a)(r-d)(r-f)} + \sqrt{s(s-b)(s-c)(s-f)}$$

essendo: $p = (a + b + e)/2$; $q = (c + d + e)/2$
 $r = (a + d + f)/2$; $s = (b + c + f)/2$

come mostrato nelle figure 4.6 e 4.7:

Gli esempi potrebbero facilmente continuare: quanti luoghi dei punti sono ben noti oltre alla retta, al piano, alle coniche ed alle quadriche. Allora, non cambia la conclusione: la determinazione delle equazioni di condizione pure è un problema da affrontare caso per caso, volta per volta. Tutto ciò non fa abbandonare completamente il metodo proposto; tuttavia lo limita notevolmente, confinandolo al caso di piccoli problemi,

dove questioni di condizionamento numerico consigliano di evitare stime di parametri, adottando, pertanto, il metodo delle equazioni di condizione pure.

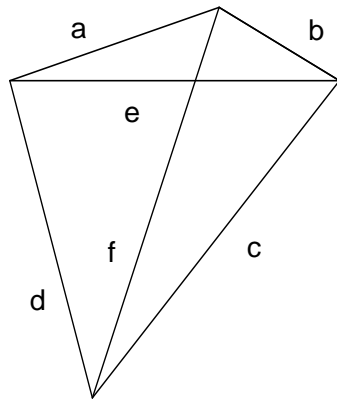


Fig. 4.6

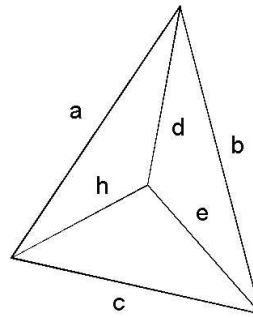


Fig. 4.7

A mo' di prova ulteriore, nella figura 4.8, una digressione sulle cubiche (studiate già da Newton e successivamente da Stirling il cui numero 78 è ben più grande di quello delle coniche), comunemente sconosciute, tranne per la parabola cubica (funzione con un flesso ed, eventualmente, un minimo ed un massimo), presenta cinque esempi di parabole: con ovale, cuspidata, nodata, puntata e pura.

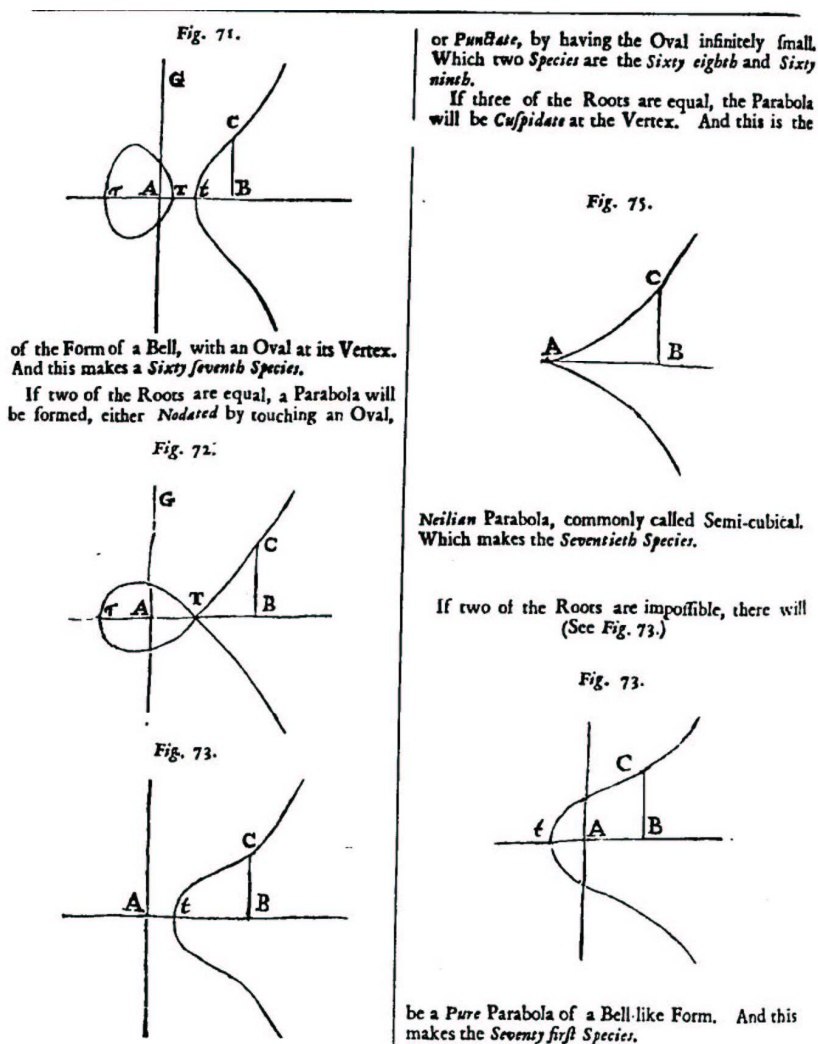


Fig. 4.8

5. EQUAZIONI DI CONDIZIONE CON PARAMETRI

Un'alternativa vantaggiosa alle equazioni di condizione pure è fornita dalle equazioni di condizione con parametri, conseguente all'introduzione nelle prime di opportuni parametri, solitamente, in numero da lasciare una sola osservabile per ogni equazione, che nel loro insieme costituiscono, a tal punto, le equazioni d'osservazione.

Si noti che più antiche espressioni denominano le stesse equazioni come equazioni delle osservazioni indirette, mentre le equazioni di condizione pure sono dette, nello stesso ambito, equazioni delle osservazioni dirette condizionate. D'altra parte l'adozione di espressioni più moderne facilita la comprensione d'assieme del metodo dei minimi quadrati e delle sue varie articolazioni e, proprio per questa ragione, dette espressioni sono adottate.

Si osservi, poi, come la scelta di operare con una sola osservabile per equazione costituisce un caso solo apparentemente particolare che, invece, come già detto, nulla toglie alla generalità del problema. Infatti tanto più generali equazioni di condizione con parametri, quanto l'inclusione di osservabili secondarie nelle usuali equazioni d'osservazione, come pure una partizione nell'insieme dei parametri fra parametri principali e parametri di servizio, esemplificano schemi differenti che possono essere ricondotti, comunque, allo schema principe delle equazioni d'osservazione.

Si badi, altresì, che anche un difetto di rango nell'insieme dei parametri, cosa che richiede l'aggiunta di opportune equazioni di vincolo o pseudo-osservazione, per sanare lo stesso, non allontana dallo schema principe prescelto. Infatti mentre l'aggiunta di equazioni di vincolo forma un insieme di equazioni di condizione con parametri vincolati (già note, altrimenti, come equazioni delle osservazioni indirette condizionate), la trasformazione delle stesse equazioni in equazioni di pseudo-osservazione sovrappesate, come sarà mostrato nel prosieguo, permette proprio l'utilizzo dello schema principe delle equazioni d'osservazione.

Un problema minimi quadrati costituito da m equazioni d'osservazione delle osservabili y , coinvolgenti n parametri x , presenta il seguente modello funzionale lineare o linearizzato:

$$y = Ax + \delta$$

dove A rappresenta la matrice dei coefficienti e δ il vettore dei termini noti delle suddette equazioni, ed il modello stocastico:

$$E(y) = \bar{y}$$

$$D(y) = C_{yy} = \sigma_0^2 Q_{yy}$$

per il quale valgono le stesse considerazioni annunciate a proposito delle equazioni di condizione pure.

Ancora le stesse considerazioni sono da farsi per quanto riguarda: la non soddisfazione, in generale, di tutte le equazioni di osservazione da parte di un campione di m osservazioni y_0 ; l'indeterminazione nella definizione di opportune variabili correzione o scarti-residui v ; l'arbitrarietà di un criterio di scelta di queste ultime variabili. Pertanto avendo stabilito l'adozione del metodo dei minimi quadrati, la sua formalizzazione matematica fa sì che una norma quadratica degli scarti-residui sia minimizzata, tenendo conto tanto del modello funzionale, quanto del modello stocastico, già definiti in precedenza:

$$\begin{aligned}\phi &= \frac{1}{2}(\hat{y} - y_0)^T Q_{y_0 y_0}^{-1} (\hat{y} - y_0) + \lambda^T (A\hat{x} + \delta - \hat{y} - y_0) = \\ &= \frac{1}{2}\hat{v}^T P\hat{v} + \lambda^T (A\hat{x} + \delta - \hat{y} - y_0) = \min\end{aligned}$$

essendo P la matrice dei pesi delle osservazioni e λ e μ un vettore ausiliario di moltiplicatori di Lagrange.

Si riportano di seguito il valore delle stime dei parametri, delle osservabili compatibili con le relazioni di condizione e delle correzioni o scarti-residui, facendo uso della nozione di matrice inversa generalizzata (definita nei paragrafi precedenti, relativamente al secondo dei due casi esposti) ed omettendo inutili passaggi algebrici elementari.

$$\begin{aligned}\hat{x} &= -(A^T P A)^{-1} (A^T P (\delta - y_0)) \\ \hat{y} &= A\hat{x} + \delta = y_0 + \hat{v} \\ \hat{v} &= \hat{y} - y_0 = A\hat{x} + \delta - y_0\end{aligned}$$

Ancora la nota legge di propagazione della varianza-covarianza permette di ricavare, facilmente, la matrice di varianza-covarianza delle stime effettuate:

$$\begin{aligned}C_{\hat{x}\hat{x}} &= \hat{\sigma}_0^2 (A^T P A)^{-1} \\ C_{\hat{y}\hat{y}} &= \hat{\sigma}_0^2 A Q_{\hat{x}\hat{x}} A^T = C_{y_0 y_0} - C_{\hat{v}\hat{v}} \\ C_{\hat{v}\hat{v}} &= C_{y_0 y_0} - C_{\hat{y}\hat{y}} = \hat{\sigma}_0^2 (P^{-1} - A Q_{\hat{x}\hat{x}} A^T) \\ C_{\hat{y}\hat{x}} &= -\hat{\sigma}_0^2 A Q_{\hat{x}\hat{x}} = C_{\hat{x}\hat{y}}^T \\ C_{\hat{x}\hat{v}} &\equiv 0 \quad ; \quad C_{\hat{y}\hat{v}} \equiv 0\end{aligned}$$

dove la stima del fattore di proporzionalità (sigma zero):

$$\hat{\sigma}_0^2 = \frac{\hat{v}^T P \hat{v}}{m - n}$$

è ottenuta imponendo, sempre direttamente, la sua correttezza. Infine tutte le stime godono e mancano delle stesse proprietà enunciate a proposito delle equazioni di condizione pure.

6. EQUAZIONI DI CONDIZIONE CON PARAMETRI VINCOLATI

Un problema minimi quadrati costituito da m equazioni d'osservazione delle osservabili y , coinvolgenti n parametri x , collegati fra loro da l equazioni di vincolo (altrimenti detto, con un'espressione più antica, equazioni delle osservazioni indirette condizionate), presenta il seguente modello funzionale lineare o linearizzato:

$$y = Ax + \delta$$

$$Hx + \eta = 0$$

dove A e H rappresentano le matrici dei coefficienti e δ ed η i vettori dei termini noti delle suddette equazioni, ed il modello stocastico:

$$E(y) = \bar{y}$$

$$D(y) = C_{yy} = \sigma_0^2 Q_{yy}$$

anche per il quale valgono le stesse considerazioni enunciate a proposito delle equazioni di condizione pure. Ancora le stesse considerazioni sono da farsi per quanto riguarda: la non soddisfazione, in generale, di tutte le equazioni d'osservazione da parte di un campione di m osservazioni y_0 ; l'indeterminazione nella definizione di opportune variabili correzione o scarti-residui v ; l'arbitrarietà di un criterio di scelta di queste ultime variabili. Pertanto avendo stabilito l'adozione del metodo dei minimi quadrati, la sua formalizzazione matematica fa sì che una norma quadratica degli scarti-residui sia minimizzata, tenendo conto tanto del modello funzionale, quanto del modello stocastico, già definiti in precedenza:

$$\begin{aligned} \phi &= \frac{1}{2} (\hat{y} - y_0)^T Q_{y_0 y_0}^{-1} (\hat{y} - y_0) + \lambda^T (A\hat{x} + \delta - \hat{v} - y_0) + \mu^T (H\hat{x} + \eta) = \\ &= \frac{1}{2} \hat{v}^T P \hat{v} + \lambda^T (A\hat{x} + \delta - \hat{v} - y_0) + \mu^T (H\hat{x} + \eta) = \min \end{aligned}$$

essendo P la matrice dei pesi delle osservazioni e λ e μ due vettori ausiliari di moltiplicatori di Lagrange.

L'elenco delle stime e delle informazioni sulla loro attendibilità non è, qui, riportato, perché il loro calcolo, quantunque possibile, risulta laborioso ed oneroso; infatti richiede la soluzione, a cascata, di più sistemi concatenati fra loro. Per queste ragioni, come già detto, si preferisce effettuare la trasformazione delle equazioni di vincolo in equazioni di pseudo-osservazione sovrappesate e risolvere un problema minimi quadrati costituito da equazioni d'osservazione e pseudo-osservazione sovrappesate, anziché d'osservazione e vincolo.

L'artificio che permette il passaggio da un problema all'altro è basato su considerazioni di carattere pratico e consiste nell'interpretare il numero zero posto al secondo membro di ogni equazione di vincolo come un'estrazione da una variabile casuale avente media e varianza nulla, cosa che fa di questa variabile casuale una pseudo-osservabile z e dello zero estratto una pseudo-osservazione z_0 .

Ancora considerazioni di ordine pratico: la necessità di calcolare un peso proporzionale all'inverso della varianza, impone di approssimare la varianza nulla con una varianza piccolissima, cosicché il peso sia sì molto grande, ma non infinito, ovvero numericamente non trattabile.

Pertanto, come già detto in precedenza, l'artificio della trasformazione delle equazioni di vincolo in equazioni di pseudo-osservazione sovrappesate permette proprio l'utilizzo dello schema principe delle equazioni d'osservazione. Inoltre lo stesso artificio ha carattere, del tutto, generale e può essere esteso a qualsiasi

generalizzazione dei modelli lineari stessi, come sarà mostrato nel prosieguo.

Un problema minimi quadrati costituito da m equazioni d'osservazione delle osservabili y , coinvolgenti n parametri x , collegati fra loro di l equazioni di pseudo-osservazioni, presenta il seguente modello funzionale lineare o linearizzato:

$$y = Ax + \delta$$

$$z = Hx + \eta = 0$$

ed il seguente modello stocastico:

$$E(y) = \bar{y} \quad ; \quad E(z) = \bar{z} = 0$$

$$D(y) = C_{yy} = \sigma_0^2 Q_{yy}$$

$$D(z) = C_{zz} = \sigma_0^2 Q_{zz} \cong 0$$

$$D(y, z) = 0$$

per il quale valgono, ancora, le stesse considerazioni enunciate in precedenza. Anche in questo caso, un campione di m osservazioni y_0 richiede di introdurre opportune variabili correzione o scarti-residui v ; le stesse variabili u devono tuttavia essere introdotte per un campione di l pseudo-osservazioni z_0 . Pertanto avendo stabilito, al solito, l'adozione del metodo dei minimi quadrati, la sua formalizzazione matematica fa sì che una norma quadratica di tutti gli scarti-residui sia minimizzata, tenendo conto tanto del modello funzionale, quanto del modello stocastico, già definiti in precedenza:

$$\phi = \frac{1}{2} \left((\hat{y} - y_0)^T Q_{y_0 y_0}^{-1} (\hat{y} - y_0) + (\hat{z} - z_0)^T Q_{z_0 z_0}^{-1} (\hat{z} - z_0) \right) +$$

$$+ \lambda^T (A\hat{x} + \delta - \hat{v} - y_0) + \mu^T (H\hat{x} + \eta - \hat{u} - z_0) =$$

$$= \frac{1}{2} (\hat{v}^T P \hat{v} + \hat{u}^T Q \hat{u}) + \lambda^T (A\hat{x} + \delta - \hat{v} - y_0) + \mu^T (H\hat{x} + \eta - \hat{u} - z_0) = \min$$

essendo P la matrice dei pesi delle osservazioni, $Q = hI$ la matrice dei pesi delle pseudo-osservazioni, dove h è un numero positivo opportunamente grande, e λ e μ due vettori ausiliari di moltiplicatori di Lagrange.

Come già detto, questo problema può essere ricondotto allo schema principe delle equazioni d'osservazione. Infatti effettuate le seguenti posizioni, che ridefiniscono, opportunamente, gli elementi del modello funzionale e del modello stocastico:

$$A = \begin{bmatrix} A \\ H \end{bmatrix} \quad ; \quad \delta = \begin{bmatrix} \delta \\ \eta \end{bmatrix} \quad ; \quad P = \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}$$

$$y_0 = \begin{bmatrix} y_0 \\ z_0 \end{bmatrix} ; \quad \hat{y} = \begin{bmatrix} \hat{y} \\ \hat{z} \end{bmatrix} ; \quad \hat{v} = \begin{bmatrix} \hat{v} \\ \hat{u} \end{bmatrix}$$

tale problema diventa, ovviamente, formalmente identico al problema minimi quadrati, già risolto, costituito da $m + l$ equazioni d'osservazione delle osservabili y e z , coinvolgenti n parametri x .

Si riportano di seguito il valore delle stime dei parametri, delle osservabili e delle pseudo-osservabili compatibili con le relazioni di condizione e delle correzioni o scarti-residui di entrambe, tenendo conto delle posizioni, che ridefiniscono gli elementi del modello funzionale e del modello stocastico, già definiti in precedenza:

$$\hat{x} = -(A^T P A + H^T Q H)^{-1} (A^T P (\delta - y_0) + H^T Q \eta)$$

$$\hat{y} = A \hat{x} + \delta = y_0 + \hat{v}$$

$$\hat{v} = \hat{y} - y_0 = A \hat{x} + \delta - y_0$$

$$\hat{z} = H \hat{x} + \eta = \hat{u} = 0$$

L'uguaglianza a zero delle stime relative alle equazioni di pseudo-osservazione dimostra il completo soddisfacimento del vincolo e discende dalle proprietà di correttezza e consistenza, godute da tutte le stime minimi quadrati.

Sempre la nota legge di propagazione della varianza-covarianza permette di ricavare, facilmente, le matrici di varianza delle stime effettuate, tenendo conto delle posizioni già definite in precedenza:

$$C_{\hat{x}\hat{x}} = \hat{\sigma}_0^2 (A^T P A + H^T Q H)^{-1}$$

$$C_{\hat{y}\hat{y}} = \hat{\sigma}_0^2 A Q_{\hat{x}\hat{x}} A^T = C_{y_0 y_0} - C_{\hat{v}\hat{v}}$$

$$C_{\hat{v}\hat{v}} = C_{y_0 y_0} - C_{\hat{y}\hat{y}} = \hat{\sigma}_0^2 (P^{-1} - A Q_{\hat{x}\hat{x}} A^T)$$

$$C_{\hat{y}\hat{x}} = \hat{\sigma}_0^2 A Q_{\hat{x}\hat{x}} = C_{\hat{x}\hat{y}}^T$$

$$C_{\hat{x}\hat{v}} \equiv 0 ; \quad C_{\hat{v}\hat{y}} \equiv 0$$

$$C_{\hat{z}\hat{z}} = C_{z_0 z_0} - C_{\hat{u}\hat{u}} = 0$$

$$C_{\hat{u}\hat{u}} = C_{z_0 z_0} - C_{\hat{z}\hat{z}} = 0$$

$$C_{\hat{x}\hat{z}} = 0$$

$$C_{\hat{x}\hat{u}} \equiv 0 ; \quad C_{\hat{y}\hat{u}} \equiv 0$$

$$C_{\hat{y}\hat{z}} = 0 \quad ; \quad C_{\hat{v}\hat{z}} \equiv 0$$

$$C_{\hat{v}\hat{u}} \equiv 0 \quad ; \quad C_{\hat{z}\hat{u}} = 0$$

dove la stima del fattore di proporzionalità (sigma zero):

$$\hat{\sigma}_0^2 = \frac{\hat{v}^T P \hat{v}}{m - n + l}$$

è ottenuta imponendo, ancora direttamente, la sua correttezza.

L'uguaglianza a zero delle matrici di varianza-covarianza, relative alle equazioni di pseudo-osservazione, completa la dimostrazione del completo soddisfacimento del vincolo e discende dal teorema di decomposizione ortogonale della varianza (come noto, infatti, il metodo dei minimi quadrati separa nelle osservazioni due termini incorrelati, una compatibile con le relazioni di condizione ed un altro residuo) e dalle disuguaglianze di Schwarz per i termini misti, quando non identicamente nulli.

Anche in questo caso, come in parte già detto, tutte le stime godono e mancano delle stesse proprietà enunciate a proposito delle equazioni di condizione pure.

Infine resta da sottolineare, come la trasformazione delle equazioni di vincolo in equazioni di pseudo-osservazione sovrappesate introduce un problema minimi quadrati più generale, dove equazioni di pseudo-osservazione di peso qualsiasi (e non di peso opportunamente grande, come per il passaggio da un problema all'altro) possono rappresentare, tra l'altro, informazioni a priori, osservabili secondarie, condizioni numeriche di regolarizzazione. D'altra parte, la soluzione di quest'ultimo problema coincide con quella del precedente, sopra riportata, tranne per l'uguaglianza a zero del valore delle stime e delle matrici di varianza-covarianza relative alle equazioni di pseudo-osservazione sovrappesate (in nessun caso, infatti, si è fatto uso dell'essere detti pesi opportunamente grandi).

Si riportano di seguito tutti gli elementi riferiti alle equazioni di pseudo-osservazione di peso qualsiasi:

$$\hat{x} = -(A^T P A + H^T Q H)^{-1} (A^T P (\delta - y_0) + H^T Q \eta)$$

$$\hat{y} = A \hat{x} + \delta = y_0 + \hat{v}$$

$$\hat{v} = \hat{y} - y_0 = A \hat{x} + \delta - y_0$$

$$\hat{z} = H \hat{x} + \eta = z_0 + \hat{u}$$

$$\hat{u} = \hat{z} - z_0 = H \hat{x} + \eta - z_0$$

$$C_{\hat{x}\hat{x}} = \hat{\sigma}_0^2 (A^T P A + H^T Q H)^{-1}$$

$$C_{\hat{y}\hat{y}} = \hat{\sigma}_0^2 A Q_{\hat{x}\hat{x}} A^T = C_{y_0 y_0} - C_{\hat{v}\hat{v}}$$

$$C_{\hat{v}\hat{v}} = C_{y_0 y_0} - C_{\hat{y}\hat{y}} = \hat{\sigma}_0^2 (P^{-1} - A Q_{\hat{x}\hat{x}} A^T)$$

$$C_{\hat{y}\hat{x}} = \hat{\sigma}_0^2 A Q_{\hat{x}\hat{x}} = C_{\hat{x}\hat{y}}^T$$

$$C_{\hat{x}\hat{y}} \equiv 0 \quad ; \quad C_{\hat{y}\hat{y}} \equiv 0 \quad ; \quad C_{\hat{x}\hat{u}} \equiv 0 \quad ; \quad C_{\hat{y}\hat{u}} \equiv 0$$

$$C_{\hat{z}\hat{z}} = \hat{\sigma}_0^2 H Q_{\hat{x}\hat{x}} H^T = C_{z_0 z_0} - C_{\hat{u}\hat{u}}$$

$$C_{\hat{u}\hat{u}} = C_{z_0 z_0} - C_{\hat{z}\hat{z}} = \hat{\sigma}_0^2 (Q^{-1} - H Q_{\hat{x}\hat{x}} H^T)$$

$$C_{\hat{x}\hat{z}} = -\hat{\sigma}_0^2 Q_{\hat{x}\hat{x}} H^T = C_{\hat{z}\hat{x}}^T$$

$$C_{\hat{y}\hat{z}} = \hat{\sigma}_0^2 A Q_{\hat{x}\hat{x}} H^T = C_{\hat{z}\hat{y}}^T$$

$$C_{\hat{v}\hat{u}} = -C_{\hat{y}\hat{z}} \quad ; \quad C_{\hat{v}\hat{z}} = 0 \quad ; \quad C_{\hat{z}\hat{u}} = 0$$

$$\hat{\sigma}_0^2 = \frac{\hat{v}^T P \hat{v} + \hat{u}^T Q \hat{u}}{m - n + l}$$

La presentazione di tutti gli elementi, riferiti alle equazioni di pseudo-osservazione di peso qualsiasi, dai vettori della soluzione alle loro matrici di varianza-covarianza ed a sigma zero, permette di sottolineare l'assoluta generalità del metodo, dove le pseudo-osservazioni di peso qualsiasi svolgono compiti anche sostanzialmente diversi dai vincoli. Infatti ogni qualvolta, per un qualsiasi motivo, sono date informazioni, a priori, e quando le stesse non possano essere considerate certe, rispetto alla variabilità delle osservazioni, in esame, trattare queste informazioni come vincoli, significa forzare ad esse la soluzione trovata, trasferendo arbitrariamente i loro errori di compatibilità sulla precisione delle osservazioni. Al contrario, il peso qualsiasi permette di avere scarti-residui anche per le suddette informazioni, conservando il rigore voluto, tanto nel modello funzionale, quanto in quello stocastico, e permettendo così un'appropriata analisi statistica dei risultati ottenuti.

7. CASI PARTICOLARI COMPLEMENTARI

I seguenti casi particolari complementari permettono di mostrare, come la soluzione a minimi quadrati di un problema con equazioni di condizione con parametri sia del tutto generale e permetta di risolvere ogni tipo di problema che possa apparentemente mostrarsi in una forma diversa. L'importanza di quest'affermazione risiede nel fatto che una soluzione diversa, da problema a problema, non costruisce bene un metodo ed apre la porta alla soluzione separata d'innumerabili esempi, uno diverso dall'altro, mentre una soluzione unica fa teoria e suggerisce una strada obbligata, insieme rigorosa e del tutto praticabile, dando così maggiore importanza al metodo costruito.

Un primo caso particolare complementare di grande interesse è costituito da equazioni di pseudo-osservazione (informazioni a priori, osservabili secondarie, condizioni numeriche di regolarizzazione), agenti ciascuna su un solo parametro, cosicché un'opportuna matrice proiettore Π può far sì che queste agiscano, apparentemente, su tutti i parametri:

$$\Pi \hat{x} = \Pi (x_0 + \hat{u})$$

$$C_{x_0 x_0} = \hat{\sigma}_0^2 \Pi Q^{-1} \Pi$$

Infatti la matrice proiettore Π è una matrice diagonale ($\dim \Pi = n$), avente l elementi unitari dove le equazioni di pseudo-osservazione agiscono, effettivamente, sui parametri ($Tr \Pi = l$) e $n-l$ elementi nulli altrove; inoltre come ogni matrice proiettore si ha: $\Pi^2 = \Pi$, mentre la sua matrice inversa generalizzata coincide con la stessa matrice proiettore, essendo ovviamente singolare.

Si riportano di seguito il valore della stima dei parametri e la corrispondente matrice di varianza-covarianza, le cui espressioni sono state, direttamente, derivate dai problemi minimi quadrati con equazioni d'osservazione e pseudo-osservazione, risolti nel paragrafo precedente:

$$\hat{x} = -(A^T P A + \Pi Q \Pi)^{-1} (A^T P (\delta - y_0) + \Pi Q x_0)$$

$$C_{\hat{x}\hat{x}} = \hat{\sigma}_0^2 (A^T P A + \Pi Q \Pi)^{-1}$$

restando, invece, invariati tutti gli altri elementi del problema in esame.

Un secondo caso particolare complementare è costituito da equazioni d'osservazione (e, eventualmente, pseudo-osservazione) con parametri, nell'insieme dei quali è stata effettuata una partizione fra parametri principali e parametri ausiliari (o di servizio):

$$\hat{y} = y_0 + \hat{v} = A\hat{x}_1 + B\hat{x}_2 + \delta$$

oppure:

$$\hat{y} = y_0 + \hat{v} = A\hat{x}_1 + B\hat{x}_2 + \delta$$

$$\hat{z} = z_0 + \hat{u} = H\hat{x}_1 + K\hat{x}_2 + \eta$$

In entrambi i casi effettuate le seguenti posizioni, che ridefiniscono, in parte, gli elementi del modello funzionale:

$$\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} ; \quad A = [A \quad B]$$

oppure:

$$\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} ; \quad A = [A \quad B] ; \quad H = [H \quad K]$$

il problema in esame è ricondotto, totalmente, ai problemi minimi quadrati con equazioni d'osservazione e pseudo-osservazione, risolti nel paragrafo precedente.

Si badi che quest'ultimo problema, quantunque elementare, fornisce la traccia per la soluzione di un terzo ed un quarto caso particolare complementare importanti, costituiti, rispettivamente, dall'inclusione di osservabili secondarie nelle equazioni d'osservazione e da più generali equazioni di condizione:

$$\hat{y} = y_0 + \hat{v} = A\hat{x}_1 + Bz_0 + \delta$$

oppure:

$$A\hat{x}_1 + By_0 + \delta - \hat{v} = 0$$

In entrambi i casi effettuate le seguenti posizioni che ridefiniscono, in parte, gli elementi del modello funzionale e del modello stocastico:

$$\hat{y} = y_0 + \hat{v} = A\hat{x}_1 + B\hat{x}_2 + \delta$$

$$\hat{z} = z_0 + \hat{u} = \hat{x}_2$$

oppure:

$$A\hat{x}_1 + B\hat{x}_2 + \delta = 0$$

$$\hat{y} = y_0 + \hat{u} = \hat{x}_2$$

i problemi in esame sono ricondotti, totalmente, al precedente caso particolare complementare e, da qui, ai problemi minimi quadrati con equazioni d'osservazione e pseudo-osservazione, risolti nel paragrafo precedente. Infatti a seguito della definizione di un numero opportuno di parametri di servizio, sono scritte tante equazioni d'osservazione (dirette) quante sono, rispettivamente, le osservabili secondarie, oppure la totalità delle osservabili. Queste equazioni d'osservazione (dirette) sono, pertanto, da aggiungersi alle altre equazioni già scritte, rispettivamente, d'osservazione in presenza di osservabili principali e pseudo-osservazione sovrappesate, se si opera con più generali equazioni di condizione.

Ciò significa che il modello stocastico, definito rispettivamente dalla matrice dei pesi:

$$P = \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \quad ; \quad P = \begin{bmatrix} Q & 0 \\ 0 & P \end{bmatrix}$$

mette in gioco, rispettivamente, tutte equazioni d'osservazione (nel primo caso, Q è il peso delle osservabili secondarie), oppure, nell'ordine, equazioni di pseudo-osservazione sovrappesate (nel secondo caso, si ha infatti: $Q = hI$) ed equazioni d'osservazione. Allora effettuate le ulteriori seguenti posizioni, che ridefiniscono, in parte, gli elementi del modello funzionale:

$$\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} \quad ; \quad A = \begin{bmatrix} A & B \\ 0 & I \end{bmatrix} \quad ; \quad \delta = \begin{bmatrix} \delta \\ 0 \end{bmatrix}$$

i problemi in esame sono ricondotti, totalmente, come già detto, ai problemi minimi quadrati con equazioni d'osservazione e pseudo-osservazione, risolti nel paragrafo precedente.

Infine occorre sottolineare, come questo modo di procedere eviti errori nella corretta definizione del modello stocastico, quali l'errata attribuzione dei pesi delle osservazioni, in presenza di osservabili secondarie o la mancata considerazione delle correlazioni, sempre in presenza di osservabili secondarie, oppure se si opera con più generali equazioni di condizione. Infatti mentre questo modo di procedere facilita notevolmente una corretta definizione del modello stocastico, i suddetti errori nella definizione dello stesso sono, purtroppo, relativamente frequenti, quando si opera per altra via.

Analogamente anche il caso di un problema minimi quadrati, costituito da sole relazioni di condizione (pure) fra osservabili, già presentato in precedenza:

$$By + \Delta = 0$$

può essere risolto, facendo debito uso delle posizioni fatte per gli ultimi casi particolari. Infatti posto:

$$B\hat{x} + \Delta = 0$$

$$\hat{y} = y_0 + \hat{u} = \hat{x}$$

il problema in esame è ricondotto, totalmente, al secondo caso particolare complementare esposto e, da qui, ai problemi minimi quadrati con equazioni d'osservazione e pseudo-osservazione, risolti nel paragrafo precedente. Infatti a seguito della definizione di un numero opportuno di parametri di servizio, sono scritte, ancora una volta, tante equazioni d'osservazione (dirette) quante sono le osservabili. Queste equazioni d'osservazione (dirette) sono, pertanto, da aggiungersi alle altre equazioni già scritte, di pseudo-osservazione sovrappesate.

Ciò significa che il modello stocastico, definito rispettivamente dalla matrice dei pesi:

$$P = \begin{bmatrix} Q & 0 \\ 0 & P \end{bmatrix}$$

mette in gioco, nell'ordine, equazioni di pseudo-osservazione sovrappesate (avendo infatti: $Q = hI$) ed equazioni d'osservazione (dirette). Allora effettuate le ulteriori seguenti posizioni, che ridefiniscono, in parte, gli elementi del modello funzionale:

$$A = \begin{bmatrix} B \\ I \end{bmatrix} ; \quad \delta = \begin{bmatrix} \Delta \\ 0 \end{bmatrix}$$

il problema in esame è ricondotto, totalmente, come già detto, ai problemi minimi quadrati con equazioni d'osservazione e pseudo-osservazione, risolti nel paragrafo precedente.

Resta da sottolineare come questo gioco di ricondurre un problema ad un altro sia, nell'ambito dei sistemi lineari, pressoché universale ed eviti, pertanto, lo studio di molteplici soluzioni, tutte abbastanza simili tra loro, ma non del tutto identiche. L'importanza di questa considerazione sta, da un punto di vista pratico, in se stessa e, da un punto di vista teorico, nell'impossibilità vera di circoscrivere una lista aperta, se questo gioco

di riconduzione non fosse stato possibile. Invece, proprio questo gioco di riconduzione detterà la strada per una soluzione praticamente computabile del problema ai minimi quadrati generalizzati cui si rinvia (a riguardo, si veda l'appendice sui modelli generalizzati).

Da ultimo si osservi come l'utilizzo all'indietro dello schema proposto permette una formulazione del problema minimi quadrati costituito da equazioni d'osservazione, coinvolgenti parametri, collegati fra loro da equazioni di vincolo, già illustrato nel paragrafo precedente, senza richiedere la non sovrapparametrizzazione dell'insieme dei parametri. Infatti la formulazione tradizionale richiede la non sovrapparametrizzazione dell'insieme dei parametri, cosa difficile da ottenere, in generale, perché i vincoli sono spesso aggiunti per sanare, proprio, la sovrapparametrizzazione dell'insieme dei parametri.

Allora la partizione dell'insieme dei parametri in due sottoinsiemi: uno che non determina difetto di rango nella relativa matrice dei coefficienti ed un secondo, la cui aggiunta al primo, determina difetto di rango nella matrice dei coefficienti complessiva, risponde positivamente alla bisogna. Pertanto effettuate le seguenti posizioni che ridefiniscono gli elementi del modello funzionale:

$$\hat{x}_1 = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} ; \quad A = [A \quad B] \quad ; \quad H = [H \quad K]$$

il sistema di equazioni di condizione con parametri vincolati diventa:

$$\hat{y} = y_0 + \hat{v} = A\hat{x}_1 + B\hat{x}_2 + \delta$$

$$H\hat{x}_1 + B\hat{x}_2 + \eta = 0$$

L'elenco delle stime e delle informazioni sulla loro attendibilità non è, anche qui, riportato, perché il loro calcolo, quantunque possibile, risulta ancora più laborioso ed oneroso; infatti richiede la soluzione, a cascata, di un maggior numero di sistemi concatenati fra loro. Per queste ragioni, come già detto, si preferisce effettuare la trasformazione delle equazioni di vincolo in equazioni di pseudo-osservazione sovrappesate e risolvere un problema minimi quadrati costituito da equazioni d'osservazione e pseudo-osservazione sovrappesate, anziché d'osservazione e vincolo, cosa che evita, fra l'altro, gli effetti della sovrapparametrizzazione dell'insieme dei parametri e, di conseguenza, indesiderate singolarità del sistema da risolvere.

8. PROBLEMI NON – LINEARI 1

La trattazione, fin qui svolta, sui problemi minimi quadrati ha riguardato, esclusivamente, l'ambito lineare di questi. D'altra parte, la maggior parte dei fenomeni e processi sono non – lineari e tale caratteristica si ripercuote, inevitabilmente, nella loro modellazione. Pertanto anche la trattazione dei problemi minimi quadrati deve potersi estendere in ambito non–lineare, fornendo modelli adatti all'analisi dei dati di questi fenomeni e processi.

La non – linearità dei problemi minimi quadrati può interessare:

- il modello funzionale;

- il modello stocastico;
- la norma.

Nel primo caso si ha, direttamente, la non – linearità delle equazioni di condizione pure, osservazione, pseudo–osservazione, vincolo, ecc.; nel secondo caso la matrice di varianza–covarianza delle osservazioni non è del tutto nota, solo a meno di una costante (ad es., il modello di Gauss–Helmert la partiziona secondo opportuni gruppi di osservazioni e considera tali sottomatrici note, ciascuna, a meno di una costante). Nel terzo caso, invece, la modifica della norma segna un distacco definitivo dai minimi quadrati, cui si fa ritorno per operare, iterativamente, in ambito lineare. Infatti in tutti e tre i casi, la non – linearità dei problemi non permette, in generale, una soluzione diretta degli stessi e, pertanto, si sceglie o, comunque, si preferisce risolverli in modo iterativo, operando, di volta in volta, in ambito lineare.

Lo schema generale per l'attuazione di detta procedura si compone dei seguenti passi:

1. inizializzazione del problema e preparazione dei dati di partenza, in base a informazioni a priori di sufficiente approssimazione;
2. avvio di un contatore per il controllo delle iterazioni;
3. effettuazione di un passaggio in ambito lineare ed ottenimento di risultati intermedi;
4. calcolo di una norma sui risultati intermedi ottenuti per il controllo delle iterazioni;
5. esecuzione di un test d'arresto del ciclo iterativo, capace di valutare in alternativa il contenimento della suddetta norma al di sotto di un'opportuna soglia prefissata, oppure il superamento da parte del contatore di un numero massimo d'iterazioni consentite;
6. arresto del ciclo iterativo ed ottenimento dei risultati definitivi, se il test d'arresto è soddisfatto; oppure aggiornamento dei dati di partenza in base ai risultati intermedi ottenuti, se il test d'arresto non è soddisfatto;
7. incremento del contatore per il controllo delle iterazioni;
8. ripetizione della procedura descritta, a partire dal passo 3.

La figura 8.1 illustra lo schema a blocchi della procedura iterativa, appena esposta.

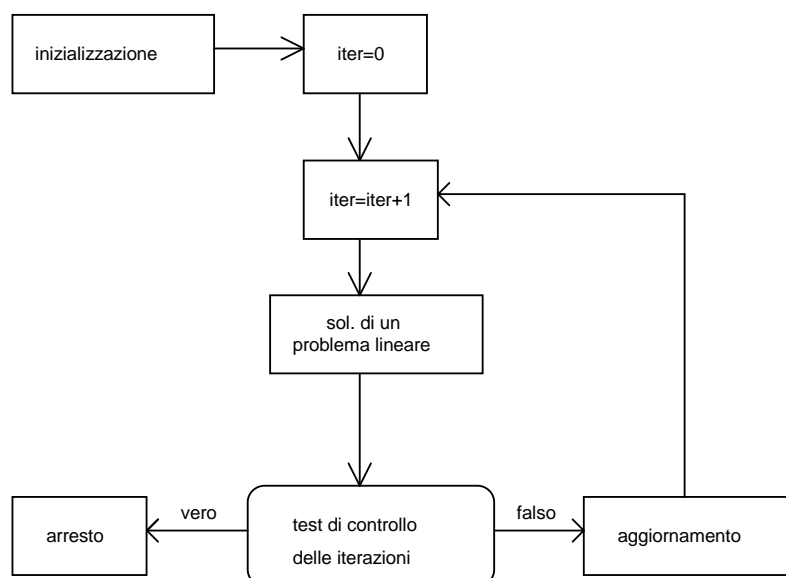


Fig. 8.1

Si osservi come tale procedura, proprio perché del tutto generale, bene si presta alla soluzione iterativa di tutti e tre i casi sopracitati di non – linearità dei problemi minimi quadrati. Infatti nessuno di essi, pur nella sua specificità, presenta anomalie tali da dover far abbandonare uno schema molto generale e, conseguentemente, molto collaudato e sicuro.

La non – linearità del modello funzionale si manifesta, come già detto, nelle equazioni d'osservazione e pseudo–osservazione. Considerazioni analoghe valgono, ovviamente, per le equazioni di condizione pure e per le equazioni di vincolo; tuttavia la trattazione a seguire restringerà l'attenzione alle sole equazioni d'osservazione e pseudo–osservazione, giudicando l'uso di equazioni di condizione pure, estremamente, raro e ricordando la possibilità di passaggio dalle equazioni di vincolo (e da più generali equazioni di condizione) alle equazioni di pseudo–osservazione. Pertanto dato un sistema d'equazioni non lineari d'osservazione e pseudo–osservazione:

$$y = F(x)$$

noto (o conosciuto per altra via) un vettore di opportuni valori approssimati dei parametri \tilde{x} , il suo sviluppo in serie di Taylor, arrestato al 1° ordine, ha form a:

$$y = F(\tilde{x}) + F_x(\tilde{x})(x - \tilde{x}) = \tilde{y} + J(\tilde{x})(x - \tilde{x})$$

essendo $J(\tilde{x})$ la matrice Jacobiano della funzione di più variabili F .

Detta matrice costituisce la matrice dei coefficienti delle nuove incognite $(\hat{x} - \tilde{x})$, mentre il vettore \tilde{y} costituisce un vettore di costanti numeriche, da aggiungersi alle osservazioni, per fornire il vettore termine noto.

Ad ogni iterazione, ottenuto il vettore delle nuove incognite $(\hat{x} - \tilde{x})$ ed aggiornato il vettore delle incognite originarie: $\hat{x} = \tilde{x} + (\hat{x} - \tilde{x})$, quest'ultimo viene considerato un nuovo vettore di più opportuni valori approssimati dei parametri, in base al quale effettuare una nuova linearizzazione, punto di partenza di una nuova iterazione.

La procedura continua, iterativamente, fino ad ottenere la convergenza del metodo alla soluzione cercata. In pratica, ad ogni iterazione, ne succede una nuova, finché le nuove incognite sono abbastanza grandi, da dare un qualche contributo, utile all'aggiornamento delle incognite originarie.

Al contrario, quando questo contributo svanisce (cioè tutte le nuove incognite sono ormai pressoché nulle), la procedura è arrestata, perché si è ottenuta la convergenza del metodo alla soluzione cercata. La procedura è arrestata, altresì, nel caso sfavorevole in cui, dopo un numero massimo d'iterazioni consentite, non si ha alcuna convergenza.

La norma calcolata sulle nuove incognite, in base alla quale giudicare l'utilità del contributo all'aggiornamento delle incognite originarie, è largamente arbitraria.

Una norma dell'estremo superiore:

$$\max_{i=1,n} = \left(\hat{x} - \tilde{x} \right)_i$$

può essere consigliabile, perché garantisce che, effettivamente, tutte le nuove incognite siano, ormai, pressoché nulle.

La procedura, appena descritta, nota come metodo di Newton–Fourier (altrimenti detto, con un'espressione più antica, degli iperpiani tangenti), può essere notevolmente semplificata e sveltita scegliendo, ad ogni iterazione dopo la prima, di ricalcolare solo il vettore \tilde{y} (ed adottando il metodo detto, dalle stesse più antiche espressioni, degli iperpiani paralleli), anziché esso e la matrice Jacobiano. Come evidente, tutto ciò accelera, ad ogni iterazione, il calcolo della soluzione, in generale, senza arrecare alcun danno alla stessa ed al solo prezzo di qualche veloce iterazione in più.

Possibili varianti al metodo di Newton-Fourier sono date da metodi più recenti, quali:

- ❑ la regolarizzazione di Tikhonov;
- ❑ l'algoritmo di Levenberg-Marquardt.

Infatti il primo metodo serve a stabilizzare la soluzione di sistemi, non troppo ben condizionati (a riguardo, si veda più oltre il paragrafo su singolarità e condizionamento), mentre il secondo metodo serve a dare peso anche alla soluzione data dai valori approssimati delle incognite, se questi possono essere considerati abbastanza soddisfacenti.

Pertanto nel primo caso, il sistema normale risulta:

$$(A^T P A + \lambda I)x + A^T P(\delta - y^0) = 0$$

mentre nel secondo caso, il sistema normale ha espressione:

$$(A^T P A + \lambda \text{diag}(A^T P A))x + A^T P(\delta - y^0) = 0$$

Per completezza si riporta anche lo sviluppo in serie di Taylor, arrestato al 4° ordine, dello stesso sistema d'equazioni non – lineari:

$$y = \tilde{y} + J(\tilde{x})(x - \tilde{x}) + (x - \tilde{x})^T H(\tilde{x})(x - \tilde{x}) + \\ + (x - \tilde{x})^T (J \otimes H(\tilde{x})) \text{vec}((x - \tilde{x})(x - \tilde{x})^T) + \\ + \text{vec}((x - \tilde{x})(x - \tilde{x})^T)^T (H \otimes H(\tilde{x})) \text{vec}((x - \tilde{x})(x - \tilde{x})^T)$$

essendo $H(\tilde{x})$ la matrice Hessiano della funzione di più variabili F , vec l'operatore capace di disporre, una dopo l'altra, le colonne di una matrice in un vettore e \otimes il simbolo indicatore del prodotto di Kronecker.

L'analisi di queste derivate ed, in particolare, di quelle di 2° ordine che danno informazioni sulla curvatura delle superfici, rappresentazione geometrica delle equazioni non – lineari, è utile per accertarsi della possibilità di effettuare una buona linearizzazione delle equazioni stesse. Infatti tale linearizzazione è possibile, geometricamente, se si ha corrispondenza biunivoca fra superfici e piani tangenti (con un esempio

cinematico, proprio della meccanica razionale, un punto materiale, mobile su una di queste superfici, ha la sua proiezione, sul corrispondente piano tangente, sempre in movimento) e, statisticamente, se la distribuzione di probabilità delle osservabili è, ivi, ben concentrata, ovvero, al di fuori della regione della corrispondenza biunivoca, la probabilità è nulla o quasi.

L'utilizzo di queste equazioni, invece, tranne per un limitatissimo numero di casi particolari (ad es., una sola equazione di 4° grado in una sola incognita; due equazioni di 2° grado, non riducibili, in due incognite; tre equazioni di 2° grado, parzialmente riducibili, per una simmetria di rotazione attorno all'asse polare, in tre incognite), trova un ostacolo insormontabile nei teoremi di Abel e Galois che affermano l'impossibilità di trovare, in generale, radici qualsiasi di equazioni di grado superiore al 3° e 4° (grado per le quali sono note, dal '500, le cosiddette soluzioni rispettivamente di Dal Ferro – Tartaglia – Cardano e Ferrari, per completezza riportate nell'immediato prosieguo).

La soluzione delle equazioni di 3° e 4° grado si determina effettuando innanzitutto una sostituzione di variabile, al fine di annullare il coefficiente dell'incognita di grado appena inferiore a quello massimo (operazione sempre possibile per ogni equazione algebrica di grado qualsiasi):

$$x = y - a_1/na_0$$

essendo x le incognite dell'equazione originaria, y quelle dell'equazione modificata ed a_0, a_1 i primi due coefficienti dell'equazione data. Dopodiché scritta un'equazione di 3° grado nella forma:

$$y^3 + py + q = 0$$

occorre esaminare il discriminante:

$$\Delta = \frac{q^2}{4} + \frac{p^3}{27}$$

dell'equazione triquadratica: $z^6 + qz^3 - p^3/27$, avendo posto: $y = z - p/3z$, e moltiplicato per z^3 .

□ Se esso è positivo, si ha una sola radice reale (essendo le altre due radici complesse coniugate):

$$y_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{\Delta}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\Delta}}$$

□ Se esso è negativo, posto: $\tan \vartheta = -2\sqrt{-\Delta}/q$, con: $0 < \vartheta < \pi/4$, si hanno tre radici reali:

4 L'uso di funzioni trigonometriche evita di dover ricorrere a numeri immaginari e complessi coniugati, essendo dimostrato che il calcolo di più radici reali non può prescindere da essi, se si effettuano solo operazioni algebriche, diversamente da quanto avviene per le equazioni di 2° grado.

$$y_1 = 2\sqrt{-\frac{p}{3}} \cos \frac{\vartheta}{3} \quad y_2 = 2\sqrt{-\frac{p}{3}} \cos \left(\frac{\vartheta}{3} + \frac{2\pi}{3} \right) \quad y_3 = 2\sqrt{-\frac{p}{3}} \cos \left(\frac{\vartheta}{3} - \frac{2\pi}{3} \right)$$

□ Se esso è nullo, si hanno due radici coincidenti reali ed una terza ancora reale:

$$y_1 = 2\sqrt{-\frac{p}{3}} \quad y_2 = y_3 = \sqrt{-\frac{p}{3}} \quad \text{tutte coincidenti, se: } p = 0$$

Invece scritta un'equazione di 4° grado nella forma:

$$y^4 + py^2 + qy + r = 0$$

occorre introdurre tre incognite di servizio u , v e w , tali che: $u + v + w = y$, ed inoltre 5:

$$u^2 + v^2 + w^2 = -\frac{p}{2} \quad u^2v^2 + v^2w^2 + w^2u^2 = \frac{p^2 - 4r}{16} \quad u^2v^2w^2 = \frac{q^2}{64}$$

Allora i quadrati delle tre incognite di servizio sono le radici dell'equazione cubica, detta risolvete di Eulero:

$$z^3 + \frac{p}{2}z^2 + \frac{p^2 - 4r}{16}z - \frac{q^2}{64} = 0$$

da cui risulta:

$$y = \sqrt{z_1} + \sqrt{z_2} + \sqrt{z_3}$$

dopo aver seguito il procedimento per la soluzione di un'equazione di 3° grado (esposto appena sopra) e prendendo i segni dei radicali, cosicché si abbia: $uvw = q/8$. Come noto, essendo un'equazione di grado pari, può avere quattro, oppure due, od addirittura nessuna radice reale e, tra le radici reali, le coincidenze possono essere tutte quattro, a due a due su quattro, solo due su quattro e due su due.

Resta da osservare come invece tutte le radici razionali delle equazioni algebriche di grado qualsiasi possano essere calcolate con la nota formula di Ruffini (di poco, precedente i teoremi di Abel e Galois) e come questa possa essere utilizzata anche in modo variazionale, cercando di minimizzare l'errore tra una radice irrazionale o trascendente e la sua approssimazione razionale, avendo scelto e fissato un certo livello d'approssimazione con cui eseguire i calcoli richiesti. Tutto ciò non vuole incentivare soluzioni dirette assurde di problemi non-lineari, ma riconoscere volentieri come studi di particolare dettaglio, se del caso, possano essere affrontati con strumenti matematici adeguati.

5 Si noti come, dati tre elementi lineari, riferiti ad un parallelepipedo, la loro somma corrisponda alla somma dei suoi tre spigoli, il loro prodotto al volume del parallelepipedo stesso e la somma delle tre coppie di prodotti distinti alla sua semi - superficie. Gli stessi termini costituiscono i tre invarianti di una matrice diagonale di dimensione e rango tre, così come somma (traccia) e prodotto (determinante) sono gli invarianti di una matrice diagonale di rango due, nonché semi - perimetro ed area di un rettangolo.

9. PROBLEMI NON – LINEARI 2

La non – lineartà del modello stocastico si manifesta, come già detto, nella matrice di varianza–covarianza delle osservazioni, non del tutto nota, solo a meno di una costante. La trattazione a seguire restringerà l'attenzione al modello di Gauss–Helmert che, avendo partizionato questa matrice in sottomatrici (detti blocchi) per opportuni gruppi di osservazioni, considera tali sottomatrici note, ciascuna a meno di una costante. Modelli più generali che coinvolgono anche elementi o blocchi extra–diagonali, fornendo informazioni sulle correlazioni fra osservazioni (o, meglio, sui parametri delle funzioni di correlazione fra gruppi di osservazioni), per l'effettiva difficoltà di stabilizzare queste stime, non sono considerati. D'altra parte, la debolezza della consistenza di stime di covarianza, anche per campioni di osservazioni abbastanza numerosi, limita l'adozione di questi modelli solo all'approccio dei processi stocastici e, a sua volta, esso, per estensione e complessità, è considerato estraneo agli scopi del presente lavoro. Allora con specifico riferimento al modello di Gauss–Helmert, la matrice di varianza–covarianza delle osservazioni ha forma:

$$C_{yy} = \sigma_0^2 (I \otimes \sigma_i^2 P_i^{-1} / \sigma_0^2)$$

dove il simbolo \otimes indica il prodotto di Kronecker fra una matrice identità, di dimensione pari al numero h di gruppi di osservazioni, stabilito dalla partizione, e la matrice (variabile per dimensione e valore degli elementi diagonali) inversa della matrice dei pesi delle osservazioni di ciascun gruppo P_i nota, come già detto, a meno di una costante σ_i^2 (sigma zero moltiplica e divide l'espressione ed è, pertanto, ininfluente). A rigore, l'utilizzo del prodotto di Kronecker, data la variabilità della seconda matrice, costituisce una leggera forzatura; d'altra parte, esso non inficia minimamente la trattazione a seguire ed aver evitato indicatori, contatori, puntatori, puntatori rovesci, ecc. la semplifica, invece, di molto, cosa che ne facilita l'esposizione. La definizione di una prima matrice di varianza–covarianza delle osservazioni, noto (o conosciuto per altra via) un vettore di opportuni valori approssimati delle costanti $\tilde{\sigma}_i^2 (i = 1, h)$, avvia la procedura cosiddetta di riproduzione dei pesi.

Ad ogni iterazione, ottenuto il vettore degli scarti–residui \hat{v} e calcolato il vettore delle stime delle costanti $\hat{\sigma}_i^2$, quest'ultimo viene considerato un nuovo vettore di più opportuni valori approssimati delle costanti, in base al quale procedere alla definizione di una nuova matrice di varianza–covarianza delle osservazioni, punto di partenza di una nuova iterazione. La procedura continua, iterativamente, fino ad ottenere la convergenza del metodo ad un punto di riproduzione. In pratica, ad ogni iterazione ne succede una nuova, finché le nuove costanti sono abbastanza diverse dalle precedenti.

Al contrario, quando le nuove costanti sono abbastanza simili alle precedenti, la procedura è arrestata, perché si è ottenuta la convergenza del metodo ad un punto di riproduzione. La procedura è arrestata, altresì, nel caso sfavorevole in cui, dopo un numero massimo d'iterazioni consentite, non si ha alcuna convergenza. La norma calcolata sulle nuove costanti, in base alla quale giudicare sulla diversificazione delle costanti fra un'iterazione ed un'altra, è largamente arbitraria. Una norma dell'estremo superiore:

$$\max_{i=1,h} (|\hat{\sigma}_i^2 - \tilde{\sigma}_i^2|)$$

è consigliabile, perché garantisce che tutte le nuove costanti siano abbastanza simili alle precedenti.

La procedura, appena descritta, nota come riproduzione dei pesi, richiede le stime delle costanti $\hat{\sigma}_i^2$, ottenute imponendo, direttamente, la loro correttezza (così come si è operato per la stima di sigma zero, con il modello di Gauss–Markov):

$$\hat{\sigma}_i^2 = \frac{(\Pi_i \hat{v})^T \Pi_i \left(I \otimes \sigma_0^2 \frac{P_j}{\sigma_j^2} \right) \Pi_i (\Pi_i \hat{v})}{v_i} \quad i = 1, h$$

essendo Π_i una matrice proiettore diversa per ogni gruppo d'osservazioni, capace di selezionare, nel vettore degli scarti–residui e nella matrice dei pesi, gli elementi appartenenti al gruppo d'osservazioni, di volta in volta, coinvolto.

Il calcolo dei denominatori v_i , numeri che determinano la correttezza delle stime delle costanti, si sviluppa come per il denominatore di sigma zero, nel modo seguente (A tale riguardo, si noti che, mentre m_i è il numero delle osservazioni appartenente al generico gruppo i , n_i è solo la quota parte dei parametri determinata dallo stesso gruppo i ; pertanto essa non è, in generale, un numero intero, né identifica nello specifico alcun parametro.):

$$\begin{aligned} v_i \sigma_i^2 &= E \left((\Pi_i \hat{v})^T \Pi_i \left(I \otimes \sigma_0^2 \frac{P_j}{\sigma_j^2} \right) \Pi_i (\Pi_i \hat{v}) \right) = Tr E \left((\Pi_i \hat{v})^T \Pi_i \left(I \otimes \sigma_0^2 \frac{P_j}{\sigma_j^2} \right) \Pi_i (\Pi_i \hat{v}) \right) \\ &= Tr \left(\Pi_i \left(I \otimes \sigma_0^2 \frac{P_j}{\sigma_j^2} \right) \Pi_i \Pi_i E(vv^T) \Pi_i \right) = \sigma_0^2 Tr \left(\Pi_i \left(I \otimes \sigma_0^2 \frac{P_j}{\sigma_j^2} \right) \Pi_i \Pi_i Q_{vv} \Pi_i \right) = \\ &= \sigma_0^2 \left(Tr \Pi_i - Tr \left(\Pi_i \left(I \otimes \frac{P_j}{\sigma_j^2} \right) \Pi_i \Pi_i A \left(A^T \left(I \otimes \frac{P_j}{\sigma_j^2} \right) A \right)^{-1} A^T \Pi_i \right) \right) = \sigma_0^2 (m_i - n_i) \end{aligned}$$

Si osservi, poi, che:

$$\frac{v_i \sigma_i^2}{\sigma_0^2} = m_i - n_i = \sum_{k=1}^{m_i} p_k q_{\hat{v}_k}$$

quantifica la ridondanza locale di tutte le osservazioni appartenenti al generico gruppo i (come noto, somma di tutte le ridondanze locali delle singole osservazioni appartenenti allo stesso gruppo) e che la somma delle ridondanze locali, estesa a tutti i gruppi, fornisce la ridondanza globale: $v = m - n$ ⁶.

Si badi, infine, che sigma zero, come già detto, moltiplica e divide l'espressione ed è, pertanto, ininfluente, cosicché la stima delle costanti, sostituendo v_i nell'espressione precedente, diventa:

³ Si chiama ridondanza locale di una generica osservazione il numero adimensionale (positivo e minore di uno) calcolato, nel caso di una matrice dei pesi diagonale, come prodotto del peso di quella osservazione per il cofattore dello scarto–residuo corrispondente:

$$v_i = p_i q_{\hat{v}_i} \quad \text{essendo:} \quad q_{\hat{v}_i} = \frac{1}{p_i} - a_i (A^T P A)^{-1} a_i^T$$

$$\hat{\sigma}_i^2 = \frac{(\Pi_i \hat{v})^T \Pi_i (I \otimes P_i) \Pi_i (\Pi_i \hat{v})}{(m_i - n_i)} \quad i = 1, h$$

La non – linearità della norma si manifesta, come già detto, nella funzione obiettivo che segna, così, un distacco definitivo dai minimi quadrati. D'altra parte, poiché, come noto, i sistemi lineari possono essere risolti 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 simplesso, 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.⁷

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 ε 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.

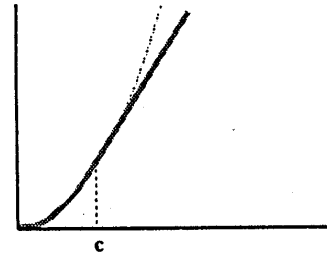
Una prima soluzione ai minimi quadrati di tutte le osservazioni (o meglio di una parte di esse, la cui bontà sia da ritenere nota o conosciuta per altra via) avvia la procedura. Gli scarti–residui di tutte le equazioni concorrono a costruire i pesi ausiliari, in base ai quali effettuare una nuova iterazione. 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 distortivi, quando si opera con i minimi quadrati (infatti, come già detto, la robustezza è una proprietà antitetica al metodo dei minimi quadrati). Le figure 9.1 e 9.2 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.

⁷ Si chiama metodo del simplesso 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

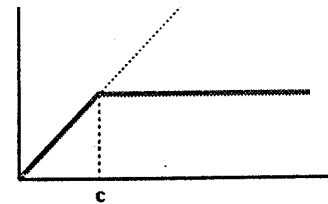
funzione obiettivo:

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



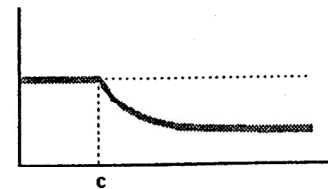
funzione d'influenza:

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



funzione dei pesi ausiliari:

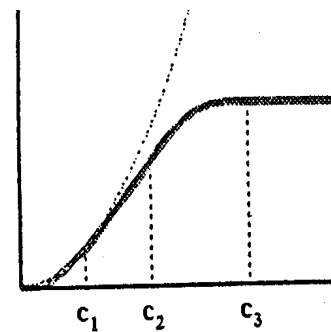
$$q(x) = \begin{cases} 1 & \|x\| < c \\ \frac{c}{\|x\|} & \|x\| \geq c \end{cases}$$



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:

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

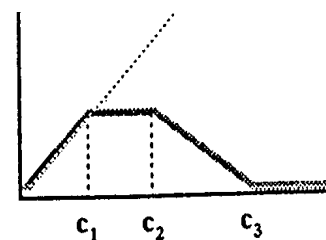


Fig. 9.1

funzione dei pesi ausiliari:

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

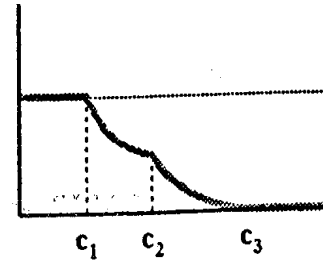


Fig. 9.2

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 ⁸, 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 ⁹, 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. Dopodichè 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. 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.

⁸ Binomio di Newton: $\binom{m}{n} = \frac{m!}{(m-n)!n!} \rightarrow m=15 \ n=5 \ \binom{m}{n} \cong 3.000; m=20 \ n=7 \ \binom{m}{n} \cong 80.000; m=30 \ n=10 \ \binom{m}{n} \cong 30.000.000.$

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.

⁹ Il numero di tali sistemi è $k = \frac{\ln \alpha}{\ln(1 - (1 - \epsilon)^p)}$, essendo α il livello di significatività del test ed ϵ la percentuale massima di dati anomali.

10. SINGOLARITA' E CONDIZIONAMENTO

E' opinione diffusa e, comunemente, accettata dai più che insiemi di dati buoni non presentino grosse difficoltà per una loro analisi, quantomeno, con la maggior parte dei metodi in uso, mentre insiemi di dati cattivi presentano notevole difficoltà per una loro analisi, in generale, con qualsiasi metodo. D'altra parte, poiché, di solito, gli insiemi di dati non sono mai troppo buoni, anche se insiemi di dati, veramente, cattivi costituiscono più controesempi che vere eccezioni, alcune precauzioni devono essere prese, per consentire una loro analisi con il metodo dei minimi quadrati, così come è necessario fare con ogni altro metodo.

Le precauzioni da prendere possono essere innumerevoli e molto diversificate tra loro; tuttavia le principali e più importanti riguardano:

- ❑ il condizionamento del sistema da risolvere;
- ❑ l'affidabilità dello schema d'osservazione;
- ❑ la robustezza degli stimatori nei confronti di dati anomali.

Un discorso sul buon condizionamento del sistema da risolvere parte dall'ovvia considerazione che esso non deve essere, globalmente o localmente, singolare ed estende le precauzioni da prendere, anche, a tutte quelle situazioni prossime alla singolarità: in tal caso, si dice che il sistema da risolvere è, globalmente o localmente, malcondizionato.

A tale proposito, un sistema di equazioni di condizione pure non è mai malcondizionato. Infatti esso ha le dimensioni del numero di relazioni di condizione e può essere singolare solo se si è scritta qualche relazione combinazione lineare delle altre, eventualmente, dimenticandone qualche altra indispensabile. Il caso descritto esemplifica un clamoroso errore d'impostazione, da evitarsi restringendo, come già detto, l'utilizzo di questo metodo al caso di piccoli problemi, dove questioni di condizionamento (numerico) consigliano di evitare stime di parametri.

Al contrario, un sistema di equazioni d'osservazione e pseudo-osservazione, benché molto raramente singolare, può essere talvolta, soprattutto localmente, malcondizionato, per la presenza di qualche parametro, maldeterminabile tramite l'insieme delle equazioni scritte. In questo caso, occorre definire una strategia d'allarme, per evidenziare il malcondizionamento, ed una strategia di ricerca, per identificare ed eliminare il parametro (i), fonte del malcondizionamento stesso.

Lo studio più rigoroso del condizionamento di un sistema si attua con la decomposizione ai valori singolari della matrice prodotto Z fra la matrice dei coefficienti A e la radice $P^{1/2}$ della matrice dei pesi (quest'ultima operazione consiste nella radice quadrata dei singoli pesi, se la matrice è diagonale, ed in una qualsiasi fattorizzazione, se essa non è diagonale):

$$P^{1/2} A = Z = WSY$$

dove S è una matrice rettangolare, avente le stesse dimensioni di A , ovunque identicamente nulla, tranne su una diagonale contenente i valori singolari, ottenuti dalla decomposizione, mentre W e Y sono due particolari matrici quadrate, dette ortonormali ($W^{-1} = W^T$; $Y^{-1} = Y^T$), aventi dimensioni compatibili con S

(ed A)¹⁰.

Nel caso la matrice A non sia una matrice di rango pieno, ovvero il sistema da risolvere sia singolare, uno o più valori singolari saranno nulli ed il loro numero sarà pari al difetto di rango.

Nel caso la matrice A sia la matrice dei coefficienti di un sistema malcondizionato, uno o più valori singolari saranno pressoché nulli. Allora una strategia d'allarme consiste nel calcolare un numero di condizione, fondato sui valori singolari, e nel valutarlo troppo prossimo a zero:

$$\varepsilon = \frac{S_{min}}{S_{max}}$$

ed una strategia di ricerca si effettua, invece, identificando tutti i valori singolari prossimi al valore singolare minimo. Un'opportuna aggiunta di osservazioni o pseudo-osservazioni, in quanto ben mirata, sarà capace di superare i problemi di condizionamento, formando un sistema ben condizionato.

D'altra parte, la decomposizione ai valori singolari è un'operazione certamente complessa e male si presta, comunque, alla soluzione di grandi sistemi; pertanto si è soliti adottare, in sua vece, una strategia alternativa d'allarme con un numero di condizione, fondato su norme di matrici, da valutare troppo prossime a zero¹¹:

¹⁰ Si chiama proiettore P in un sottospazio Y di uno spazio vettoriale X , una matrice quadrata e simmetrica, tale che il vettore $y = Px$, con $x \in X$, sia la componente del vettore x in Y . Una proprietà fondamentale di tutti i proiettori è $PP = P$, mentre, in generale, non esiste P^{-1} , tranne nel caso in cui il proiettore coincida con l'identità ($P = I$).

Si chiama rappresentazione spettrale di una matrice C quadrata, simmetrica e definita positiva la sommatoria:

$$C = \sum_{i=1}^n \lambda_i P_i = \sum_{i=1}^n \lambda_i u_i u_i^T = U \Lambda U^T$$

dove P_i è il proiettore di un qualsiasi vettore sull'iperspazio (insieme degli autovettori corrispondenti ad uno stesso autovalore). Si verifica, di seguito, la sopraccitata proprietà dei proiettori:

$$P_i^2 = (u_i u_i^T)(u_i u_i^T) = u_i (u_i^T u_i) u_i^T = u_i u_i^T = P_i$$

dove ogni u_i è un autovettore corrispondente all'autovalore λ_i .

La decomposizione ai valori singolari è l'estensione della precedente operazione di rappresentazione spettrale, ottenuta, per generalizzazione dell'argomento, dalle matrici quadrate alle matrici rettangolari.

⁷ Per semplicità, si considerino dapprima separatamente gli effetti prodotti dalle variazioni sul termine noto b e sulla matrice A , non singolare, di un generico sistema di equazioni d'osservazione.

$$Ax = b; \quad A(x + \delta x) = b + \delta b$$

Sottraendo le equazioni si ricava:

$$A\delta x = \delta b \\ \delta x = A^{-1}\delta b$$

ed applicando le definizioni di norma naturale di matrice, indicata con $\| \cdot \|$ una delle norme appartenenti a questo tipo, si ottengono le seguenti relazioni:

$$\|\delta x\| = \|A^{-1}\delta b\| \leq \|A^{-1}\| \cdot \|\delta b\|$$

$$\|b\| = \|Ax\| \leq \|A\| \cdot \|x\|$$

$$\frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|}$$

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\delta b\|}{\|b\|}$$

Si consideri, di seguito, il caso in cui x e $x + \delta x$ siano le soluzioni dei seguenti sistemi:

$$Ax = b \quad ; \quad (A + \delta A)(x + \delta x) = b$$

A partire dalla sottrazione dei due sistemi, si ottiene:

$$0 = A\delta x + \delta A(x + \delta x)$$

$$\delta x = -A^{-1} \delta A(x + \delta x)$$

e, in base a considerazioni già utilizzate in precedenza, si ricavano le seguenti relazioni:

$$\|\delta x\| = \|A^{-1}\delta A(x + \delta x)\| \leq \|A^{-1}\| \cdot \|\delta A\| \cdot \|x + \delta x\|$$

$$\frac{\|\delta x\|}{\|x + \delta x\|} \cong \frac{\|\delta x\|}{\|x\|} \leq \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\delta A\|}{\|A\|}$$

Infine, qualora si consideri l'effetto congiunto delle due perturbazioni, si hanno i seguenti sistemi:

$$Ax = b \quad ; \quad (A + \delta A)(x + \delta x) = b + \delta b$$

Sottraendo le equazioni si ricava:

$$\delta A + A\delta x + \delta A\delta x = \delta A \cdot x + A(I + A^{-1}\delta A)\delta x = \delta b$$

$$\delta x = A^{-1}(I + A^{-1}\delta A)^{-1}(\delta b - \delta A \cdot x)$$

Ed applicando le definizioni di norma naturale di matrice, si ottengono le seguenti relazioni:

$$\|\delta x\| = \|A^{-1}(I + A^{-1}\delta A)^{-1}(\delta b + \delta A \cdot x)\|$$

$$\|\delta x\| \leq \|A^{-1}\| \cdot \|(I + A^{-1}\delta A)^{-1}\| \cdot (\|\delta b\| + \|\delta A\| \cdot \|x\|)$$

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\| \cdot \|A^{-1}\| \cdot \|(I + A^{-1}\delta A)^{-1}\| \cdot \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|A\|} \frac{1}{\|x\|} \right)$$

ovvero, tenuto conto che:

$$\kappa_{\infty} = \frac{1}{\|C\|_{\infty} \cdot \|C^{-1}\|_{\infty}}$$

essendo C la matrice normale, C^{-1} la sua matrice inversa e $\|\cdot\|_{\infty}$ la norma dell'estremo superiore:

$$\|Q\|_{\infty} = \max_{i=1,n} \left(\sum_{j=1}^n |q_{ij}| \right)$$

Dopodiché una strategia alternativa di ricerca si effettua identificando, nella matrice dei coefficienti di correlazione dei parametri, tutti quei coefficienti il cui valore assoluto è, relativamente, prossimo ad uno:

$$R = (I * C^{-1})^{-1/2} C^{-1} (I * C^{-1})^{-1/2}$$

essendo C^{-1} la matrice inversa della matrice normale, I una matrice identità ed il simbolo $*$ indica il prodotto di Hadamard.

Il raggruppamento in un blocco di un insieme relazionato di coefficienti di correlazione dei parametri il cui valore assoluto è, per tutti, relativamente, prossimo ad uno, aiuta a risolvere il problema delle correlazioni multiple e, in ogni caso, un'opportuna e ben mirata aggiunta di osservazioni e pseudo-osservazioni permette di superare i problemi di condizionamento.

Nei casi più complessi, è consigliabile porre vincoli a zero (ovvero pseudo-osservazioni sovrappesate) per tutti i parametri del blocco malcondizionante, liberandone, uno alla volta, sequenzialmente, in avanti, finché non si ha malcondizionamento del sistema da risolvere ed, all'indietro, non appena questo si determina

$$\frac{1}{\|A\| \cdot \|x\|} \leq \frac{1}{\|b\|} \quad \text{e} \quad \|(I + A^{-1} \delta A)^{-1}\| \leq \frac{1}{(1 - \|A^{-1}\| \cdot \|\delta A\|)}$$

si ottiene:

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\| \cdot \|A^{-1}\| \cdot \frac{\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}}{1 - \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\delta A\|}{\|A\|}}$$

Come indice di condizionamento è stato perciò definito, in letteratura, il numero:

$$\mu(A) = \|A\| \cdot \|A^{-1}\|$$

tale definizione è stata introdotta, per matrici definite positive, da Touring nella norma di Frobenius e Van Neumann nella norma 2.

(Questa nota riprende quanto esposto in: Bellone T.. (2001): Minimi quadrati: qualità ed inconvenienti. Esempi in ambito topografico. In: Rivista dell'Agenzia del Territorio, n. 1/01.).

(l'ordine di selezione dei parametri e la scelta di quello malcondizionante, fra quelli divenuti fortemente correlati fra loro, è largamente arbitraria), con l'obiettivo di garantire, sempre, il buon condizionamento.

Per quanto riguarda, infine, l'indicazione di valori di soglia, pur nell'oggettiva difficoltà di fornire valori numerici di validità universale, un numero di condizione inferiore a $10^{-3} \div 10^{-5}$ ed un coefficiente di correlazione superiore, in valore assoluto, a $0.7 \div 0.8$ possono considerarsi valori di soglia di uso abbastanza generale.

Un'alternativa possibile è data dal seguente modo di procedere¹²:

- si aggiungano, al sistema dato, tante nuove pseudo-osservazioni (di peso comparabile con le osservazioni del suddetto sistema) ciascuna costituita da uno dei parametri del sistema stesso;

$$\begin{bmatrix} A \\ I \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{u} \end{bmatrix} + \begin{bmatrix} \delta - y_0 \\ -x_0 \end{bmatrix} = \begin{bmatrix} \hat{v} \\ \hat{u} \end{bmatrix} \quad ; \quad \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}$$

- con metodo dei minimi quadrati, si calcoli la soluzione di questo sistema esteso, dove si ha anche la matrice di covarianza degli scarti residui delle nuove pseudo-osservazioni (sui parametri):

$$C_{\hat{u}\hat{u}} = C_{x_0x_0} - C_{\hat{x}\hat{x}} = \hat{\sigma}_0^2 (Q^{-1} - Q_{\hat{x}\hat{x}})$$

- e si analizzi il vettore delle corrispondenti ridondanze locali:

$$v_i = \text{diag}(Q^{1/2} Q_{\hat{u}\hat{u}}^{-1} Q^{1/2})_i = \text{diag}(Q^{1/2} (Q^{-1} - Q_{\hat{x}\hat{x}}) Q^{1/2})_i = 1 - q_i \text{diag}(Q_{\hat{x}\hat{x}})_i$$

- poiché, come noto, tutte le ridondanze locali hanno valori compresi fra zero ed uno, raggiungendo il valore zero, quando un'osservazione (o una pseudo-osservazione) è indispensabile (anche valori troppo prossimi a zero sono indice di criticità), ed il valore uno, quando la stessa è completamente superflua:

$$0 \leq 0.20 \div 0.25 \leq v_i = 1 - q_i \text{diag}(Q_{\hat{x}\hat{x}})_i \leq 1$$

- se qualcuna di queste ridondanze locali è nulla (o troppo piccola), allora il corrispondente parametro è determinato esclusivamente (o quasi esclusivamente) dalla nuova pseudo-osservazione sul parametro stesso;
- in questo caso, è proprio necessario aggiungere nuove osservazioni al sistema, per sanare la singolarità o il malcondizionamento generati, nel sistema dato, da questo parametro (come nel paragrafo successivo sull'affidabilità delle osservazioni, sarà proprio necessario aggiungere nuove osservazioni, per sanare la non-affidabilità di un'osservazione risultata tale).

¹² Per meglio seguire questa parte conclusiva sulla singolarità ed il malcondizionamento, si consiglia di prendere in esame anche il paragrafo successivo sull'affidabilità e la robustezza, in quanto il modo di procedere proposto fa riferimento alla nozione di ridondanza locale e ne fa uso esattamente come esposto per lo studio dell'affidabilità delle osservazioni.

11. AFFIDABILITA' E ROBUSTEZZA

Uno schema d'osservazione si dice affidabile, quando è in grado di identificare uno o più dati anomali nell'insieme delle osservazioni. Ciò significa che la presenza di dati anomali, per quanto abbia, sempre, effetti distorcenti sulle stime (causa la mancanza della robustezza, come già detto, proprietà antitetica al metodo dei minimi quadrati), è grazie alla ridondanza globale e locale dello schema d'osservazione, comunque, evidenziata (cioè si sa che i dati anomali sono presenti) e localizzata (cioè si sa dove i dati anomali sono accaduti).

Con preciso riferimento ad un sistema di equazioni d'osservazione e pseudo-osservazione (ma sarebbe, certamente, possibile fare riferimento anche ad un sistema di equazioni di condizione pure, anche se quanto segue deve, allora, essere formulato in modo diverso), si considerino i due seguenti schemi delle osservazioni:

- I. un insieme d'osservazioni (e/o pseudo-osservazioni) y_{02} determina, univocamente, un insieme di parametri \hat{x}_2 , la cui dimensione è, ovviamente, identica a quella del suddetto insieme di osservazioni;
- II. due insiemi d'osservazioni (e/o di pseudo-osservazioni) y_{02} e y_{03} determinano, ciascuno, univocamente, un insieme di parametri \hat{x}_2 , la cui dimensione è, ancora ovviamente, identica a quella di ciascuno dei suddetti insiemi di osservazioni.

Pertanto il primo esempio è costituito dal seguente sistema di equazioni d'osservazione (e/o pseudo-osservazione) cui è associata la matrice dei pesi sotto riportata:

$$\begin{bmatrix} A_1 & 0 \\ A_2 & B \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} + \begin{bmatrix} \delta_1 - y_{01} \\ \delta_2 - y_{02} \end{bmatrix} = \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \end{bmatrix} \quad ; \quad \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}$$

Detto sistema, ricordando l'identità fra le dimensioni dei vettori y_{02} e \hat{x}_2 , è direttamente risolubile in \hat{x}_2 , annullando il vettore degli scarti-residui \hat{v}_2 , in quanto questa scelta minimizza, ovviamente, la norma quadratica degli stessi:

$$\hat{v}_2 \equiv 0 \quad ; \quad \hat{x}_2 = -B^{-1}((\delta_2 - y_{02}) + A_2 \hat{x}_1)$$

Tutto ciò determina la completa identità fra osservazioni e stime delle osservabili compatibili con le relazioni di condizione e, per il teorema di decomposizione ortogonale della varianza, la nullità della matrice di varianza-covarianza degli scarti-residui \hat{v}_2 :

$$C_{\hat{v}_2 \hat{v}_2} \equiv 0$$

Allora un qualsiasi dato anomalo, inserito nell'insieme delle osservazioni (e/o pseudo-osservazioni) y_{02} , distorce le stime dei parametri \hat{x}_2 , lasciando sempre nulli gli scarti-residui \hat{v}_2 con precisione infinita, essendo sempre nulla, anche, la loro matrice di varianza-covarianza.

Analogamente il secondo esempio è costituito dal seguente sistema di equazioni d'osservazione (e/o pseudo-osservazione) cui è associata la matrice dei pesi sotto riportata:

$$\begin{bmatrix} A_1 & 0 \\ A_2 & B \\ A_3 & C \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} + \begin{bmatrix} \delta_1 - y_{01} \\ \delta_2 - y_{02} \\ \delta_3 - y_{03} \end{bmatrix} = \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \hat{v}_3 \end{bmatrix} \quad ; \quad \begin{bmatrix} P & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & R \end{bmatrix}$$

Detto sistema, ricordando l'identità fra le dimensioni dei vettori y_{02} e y_{03} e \hat{x}_2 , è due volte direttamente risolubile in \hat{x}_2 , per cui la media ponderata dei valori ottenuti in modo indipendente, diventa il valore più probabile, comparando, tramite pesi opportuni, a meno del segno, i vettori degli scarti-residui \hat{v}_2 e \hat{v}_3 , in quanto questa scelta minimizza, ovviamente, la norma quadratica degli stessi:

$$\hat{v}'_2 = H\hat{v}_2 \equiv -K\hat{v}_3 = -\hat{v}'_3; \quad H \text{ e } K \text{ matrici di pesi (opportuni)}$$

$$\begin{aligned} \hat{x}_2 &= -(B^TQB + C^TRC)^{-1} (B^TQ((\delta_2 - y_{02}) + A_2\hat{x}_1) + C^TR((\delta_3 - y_{03}) + A_3\hat{x}_1)) \\ \hat{v}_2 &= B(B^TQB + C^TRC)^{-1} C^TRC (B^{-1}((\delta_2 - y_{02}) + A_2\hat{x}_1) - C^{-1}((\delta_3 - y_{03}) + A_3\hat{x}_1)) = -H^{-1}K\hat{v}_3 \\ \hat{v}_3 &= C(B^TQB + C^TRC)^{-1} B^TQB (B^{-1}((\delta_2 - y_{02}) + A_2\hat{x}_1) - C^{-1}((\delta_3 - y_{03}) + A_3\hat{x}_1)) = -K^{-1}H\hat{v}_2 \end{aligned}$$

Tutto ciò determina l'intercambiabilità (a meno del segno) delle correzioni da apportare alle osservazioni, per ottenere stime delle osservabili compatibili con le relazioni di condizione, e questa intercambiabilità, tenuto conto del teorema di decomposizione ortogonale della varianza, si ha, anche, nelle matrici di varianza-covarianza degli scarti-residui \hat{v}_2 e \hat{v}_3 :

$$\begin{aligned} C_{\hat{v}_2\hat{v}_2} &= HC_{\hat{v}_2\hat{v}_2} H^T \equiv KC_{\hat{v}_3\hat{v}_3} K^T = C_{\hat{v}_3\hat{v}_3} \\ C_{\hat{v}_2\hat{v}_2} &= H^{-1}KC_{\hat{v}_3\hat{v}_3} K^T (H^{-1})^T \quad ; \quad C_{\hat{v}_3\hat{v}_3} = K^{-1}HC_{\hat{v}_2\hat{v}_2} H^T (K^{-1})^T \end{aligned}$$

Allora un qualsiasi dato anomalo, inserito nell'insieme delle osservazioni (e/o pseudo-osservazioni) y_{02} , y_{03} , distorce le stime dei parametri \hat{x}_2 , lasciando sempre intercambiabili gli scarti-residui \hat{v}_2 e \hat{v}_3 con precisione comparabile, tramite pesi opportuni, essendo sempre intercambiabili fra loro, anche, le rispettive matrici di varianza-covarianza.

Tutto ciò diventa, particolarmente, evidente quando si ha la completa identità delle matrici dei coefficienti e delle matrici dei pesi (ovvero: $B \equiv C$ e $Q \equiv R$); in tal caso, l'intercambiabilità dei vettori degli scarti-residui e delle rispettive matrici di varianza-covarianza diventa completa identità (a meno del segno, per quanto

riguarda i vettori) degli stessi:

$$\hat{v}_2 \equiv -\hat{v}_3 \quad ; \quad C_{\hat{v}_2\hat{v}_2} \equiv C_{\hat{v}_3\hat{v}_3}$$

Infatti essendo unitari i sopraccitati opportuni pesi che permettono la comparazione (ovvero: $H = K = I$), la stima dei parametri \hat{x}_2 è la media (aritmetica) delle due determinazioni separate e gli scarti-residui \hat{v}_2 e \hat{v}_3 sono sempre uguali ed opposti:

$$\hat{y} = -\frac{1}{2}B^{-1}((\delta_2 - y_{02}) + A_2\hat{x}_1 + (\delta_3 - y_{03}) + A_3\hat{x}_1)$$

$$\hat{v}_2 = \frac{1}{2}(((\delta_2 - y_{02}) + A_2\hat{x}_1) - ((\delta_3 - y_{03}) + A_3\hat{x}_1)) \equiv -\hat{v}_3$$

$$\hat{v}_3 = -\frac{1}{2}(((\delta_2 - y_{02}) + A_2\hat{x}_1) - ((\delta_3 - y_{03}) + A_3\hat{x}_1)) \equiv -\hat{v}_2$$

espressioni, del resto, facilmente ricavabili, particolarizzando le espressioni generali sopra riportate.

I due esempi insegnano che lo schema d'osservazione è affidabile, solo quando è elevata la ridondanza globale e locale. Pertanto globalmente, il numero d'osservazioni e pseudo-osservazioni è bene sia, quantomeno, tre volte il numero dei parametri, cosicché la ridondanza globale, in termini relativi, abbia, almeno, valore due:

$$v = \frac{m+l-n}{n} \geq 2$$

Inoltre localmente, essendo difficile individuare accanto ad un sottoinsieme d'osservazioni e pseudo-osservazioni un sottoinsieme di parametri, si considera comunemente affidabile un'osservazione (o pseudo-osservazione) la cui ridondanza locale v_i abbia, almeno, valore un quinto o un quarto dell'unità:

$$0 \leq 0.20 \div 0.25 \leq v_i = 1 - p_i a_i Q_{\hat{x}\hat{x}} a_i^T \leq 1$$

Infatti come in parte già detto, la ridondanza locale di ciascuna osservazione (o pseudo-osservazione) è fornita dal corrispondente elemento diagonale principale della matrice prodotto della matrice dei cofattori degli scarti-residui preceduta e seguita dalla radice pari della matrice dei pesi (quest'ultima operazione consiste nella radice quadrata dei singoli pesi, se la matrice è diagonale, mentre trattasi d'operazione ben più complessa, se essa non è diagonale):

$$S = P^{1/2} Q_{\hat{v}\hat{v}} P^{1/2} = P^{1/2} (P^{-1} - A Q_{\hat{x}\hat{x}} A^T) P^{1/2}$$

Allora come noto, per il teorema di decomposizione ortogonale della varianza, la ridondanza locale ha valore

zero, quando un'osservazione (o pseudo-osservazione) è indispensabile, mentre ha valore limite uno, quando la stessa è, proprio, del tutto superflua. Essendo, ovviamente, impossibile avere sempre il valore uno, valori superiori ad un quinto o un quarto dell'unità si considerano, comunemente, al di sopra di una ragionevole soglia di sicurezza.

Un riesame, ancora più approfondito, della questione dell'affidabilità conduce alla definizione di affidabilità interna ed affidabilità esterna.

Si chiama affidabilità interna di una generica osservazione (e/o pseudo-osservazione) la quantità che rimane nel corrispondente scarto-residuo, a seguito dell'immissione di un effetto perturbativo.

Questa è misurabile tramite l'espressione:

$$\nabla(\delta_i - y_{0i}) = \frac{\tau \sigma_{y_{0i}}}{\sqrt{V_i}}$$

essendo:

$\sigma_{y_{0i}}$ lo sqm dell'osservazione y_{0i} ;

V_i la corrispondente ridondanza locale;

τ l'ascissa corrispondente al valore della curva di potenza (per la distribuzione di probabilità della variabile casuale τ di Thompson), scelto un livello di significatività ed una potenza del test.

Si chiama, invece, affidabilità esterna di una generica osservazione (e/o pseudo-osservazione) la quantità che fluisce nella stima di un certo parametro, a seguito dell'immissione dello stesso effetto perturbativo.

Questa è misurabile tramite l'espressione:

$$\nabla x_j = -\left((A^T P A)_j^{-1}\right)^T A^T P e_i \nabla(\delta_i - y_{0i}) \quad \forall j$$

essendo: e_i , il versore unitario diretto secondo la componente i -esima del vettore delle osservazioni (e/o pseudo-osservazioni).

Infatti data la decomposizione di una generica osservazione (e/o pseudo-osservazione) y_{0i} in un valore compatibile con le relazioni di condizione \hat{y}_i ed una discrepanza (uguale ed opposta alla correzione o scarto-residuo \hat{v}_i):

$$y_{0i} = \hat{y}_i - \hat{v}_i$$

consegue, per il teorema di decomposizione ortogonale della varianza, la decomposizione:

$$\sigma_{y_{0i}}^2 = \sigma_{\hat{y}_i}^2 + \sigma_{\hat{v}_i}^2$$

dove:

$$\sigma_{y_{0_i}}^2 = \frac{\sigma_0^2}{p_i}$$

$$\sigma_{\hat{y}_i}^2 = \frac{\sigma_0^2(1-v_i)}{p_i}$$

$$\sigma_{\hat{v}_i}^2 = \frac{\sigma_0^2 v_i}{p_i}$$

essendo, rispettivamente, p_i il peso e v_i la ridondanza locale dell'osservazione (e/o pseudo-osservazione) in esame. La standardizzazione di quest'ultima decomposizione porta alle seguenti identità:

$$z_i = \frac{y_{0_i}}{\sigma_{y_{0_i}}} = \frac{\hat{y}_i}{\sigma_{\hat{y}_i}} \cdot \frac{\sigma_{\hat{y}_i}}{\sigma_{y_{0_i}}} - \frac{\hat{v}_i}{\sigma_{\hat{v}_i}} \cdot \frac{\sigma_{\hat{v}_i}}{\sigma_{y_{0_i}}} = t\sqrt{1-v_i} - \tau\sqrt{v_i}$$

essendo z una grandezza variabile distribuita come la variabile casuale normale, t una grandezza variabile distribuita come la variabile casuale t di Student e τ una grandezza variabile distribuita come la variabile casuale τ di Thompson.

I valori teorici di dette variabili casuali sono disponibili dalle tavole di distribuzione della probabilità e rappresentano, ciascuno, l'ascissa corrispondente al valore della rispettiva curva di potenza, scelto un livello di significatività ed una potenza del test.

Allora essendo:

$$\hat{x} = -(A^T P A)^{-1} A^T P (\delta - y_0)$$

$$\hat{v} = A \hat{x} + \delta - y_0 = (I - A(A^T P A)^{-1} A^T P)(\delta - y_0)$$

una perturbazione isolata nell'osservazione (e/o pseudo-osservazione) in esame determina una perturbazione indotta, tanto nelle stime dei singoli parametri, quanto in quella dello scarto-residuo corrispondente:

$$\nabla \hat{x}_j = -\left((A^T P A)_j^{-1}\right)^T A^T P e_i \nabla(\delta_i - y_{0_i}) \quad \forall j$$

$$\nabla \hat{v}_i = \hat{v}_i = v_i \nabla(\delta_i - y_{0_i})$$

essendo e_i , il versore unitario diretto secondo la componente i -esima del vettore delle osservazioni (e/o pseudo-osservazioni).

La prima delle due relazioni definisce l'affidabilità esterna, mentre esplicitando, opportunamente, la seconda (tenuto conto del secondo addendo delle ultime due identità, definite in precedenza, ovvero:

$$\hat{v}_i = \tau \sigma_{y_{0_i}} \sqrt{v_i}), \text{ si ottiene:}$$

$$\nabla(\delta_i - y_{0_i}) = \frac{\hat{v}_i}{v_i} = \frac{\tau\sigma_{y_{0_i}}}{\sqrt{v_i}}$$

che costituisce la definizione di affidabilità interna.

Infine per quanto riguarda gli ovvi vantaggi dell'uso di stimatori robusti nei confronti di dati anomali, se uno schema d'osservazione affidabile identifica, cioè evidenzia e localizza i dati anomali, l'uso di stimatori robusti evita effetti distortivi sulle stime come accade, invece, quando si opera con i minimi quadrati (infatti, come già detto, la robustezza è una proprietà antitetica al metodo dei minimi quadrati).

APPENDICE A – RICERCA DI ERRORI GROSSOLANI¹³

A.1 I dati anomali ed il loro trattamento

La presenza di errori grossolani entro gli insiemi di dati misurati costituisce da sempre una delle principali preoccupazioni per chi opera nel campo del trattamento dei dati. E' noto infatti che anche un solo dato anomalo non individuato, può rovinare completamente un insieme di osservazioni, alterando i risultati dell'analisi del fenomeno e la precisione di stima dei parametri che lo descrivono. Conviene tuttavia analizzare, più a fondo, come nasce la definizione intuitiva che lo interpreta come un dato anomalo. La reiezione soggettiva, storicamente il primo metodo di valutazione degli errori, si basa sulla percezione di un comportamento ben definito del grosso dei dati, rispetto cui una minoranza sembra deviare alquanto.

Il punto di partenza è la definizione del grosso dei dati, cioè del comportamento delle osservazioni buone. L'uso di metodi quantitativi, cioè di un modello matematico che spieghi le osservabili in funzioni di parametri di varia natura (fisici, geometrici) e la scelta di un metodo di stima che interpreti la variabilità dei dati osservati sono l'esatto corrispondente della definizione (a priori) del comportamento delle osservazioni buone. Pertanto il dato anomalo è qualcosa di estraneo ad un modello prefissato, definito dal tipo di stimatore e dalla spiegazione scelti per il fenomeno: la sua definizione è relativa, perché è legata alla scelta dei modelli funzionale e stocastico e del criterio di stima. Posto che il modello funzionale adottato sia già consolidato, cioè che sia adeguata la comprensione e la parametrizzazione del fenomeno, il problema dell'affidabilità di un insieme di osservazioni, cioè della sua capacità di evidenziare gli errori, si pone anzitutto a livello del metodo di stima e secondariamente dello schema di misura.

Quanto più il metodo di stima riesce a cogliere la porzione dei dati coerenti con il modello, accettando invece scarti, anche elevati, laddove qualche valore si stacca dal resto in modo evidente, tanto più esso può essere definito robusto nei confronti dei dati anomali. Sotto questo punto di vista, il tradizionale criterio di stima dei minimi quadrati si presenta strutturalmente debole: infatti esso non opera scelte preferenziali fra i dati, ma accetta il contributo (definito a priori tramite il peso) di tutte le osservazioni. Inoltre ogni scarto è combinazione lineare (in generale) di tutti gli errori presenti nelle misure, cosicché si ha una difficoltà, ancora maggiore, nel tentare di risalire dai residui agli errori.

A tal fine, occorre introdurre l'ipotesi di distribuzione normale delle osservazioni ed usare gli strumenti della verifica probabilistica. Pur con una serie di limitazioni ed ipotesi semplificative che talvolta pregiudicano la

¹³ Questa appendice riprende ed aggiorna quanto esposto in: Forlani G., Mussio L. (1990): Affidabilità e ricerca di errori grossolani. In: F. Crosilla, L. Mussio (Ed's) *Dall'analitico al digitale: nuovi sviluppi della fotogrammetria applicata all'ingegneria*. CISM, Udine, p. 105–138; a sua volta, liberamente tratta da: G. Forlani (1990): Metodi robusti di stima in geodesia e fotogrammetria. In: *Ricerche di Geodesia Topografia e Fotogrammetria*, n.8, CLUP, Milano, p. 124–311.

funzionalità del metodo, si è in grado di intervenire a livello sia operativo sia progettuale. Nel primo caso, esiste una serie di test statistici, per segnalare la presenza di errori grossolani, in base ad ipotesi più o meno complesse. Nel secondo caso, è possibile, per un dato schema di misura, analizzare i punti di debolezza dello stesso e valutare (a priori) l'entità del minimo errore grossolano evidenziabile (ovvero l'affidabilità interna), come pure definire, in tale errore, la conseguente distorsione delle stime dei parametri (ovvero l'affidabilità esterna).

In alternativa ai minimi quadrati, sono state proposte procedure, cosiddette robuste, elaborate per tener conto di piccole deviazioni alle ipotesi distribuzionali delle osservazioni, quali ad esempio la presenza di errori grossolani. Scopo di questi stimatori è quello di limitare la distorsione delle stime in presenza di dati anomali, senza eccessive perdite di efficienza nel caso (favorevole) in cui non vi siano contaminazioni dei dati.

Storicamente il primo fra questi criteri, addirittura precedente quello dei minimi quadrati, è quello di minimo della somma dei moduli dei residui. Esso ha la proprietà di interpolare esattamente tanti dati quanti sono i parametri da stimare, fra tutti i valori osservati e, nel caso monodimensionale, porta alla scelta della mediana come centro delle osservazioni (in contrapposizione alla media che deriva dai minimi quadrati).

Per questo e per gli altri stimatori robusti, non esiste ancora una chiara conoscenza delle distribuzioni campionarie; pertanto non è possibile una stima della precisione, né una verifica di ipotesi. Allora è necessario integrare tali metodi con il criterio di stima dei minimi quadrati, per sfruttare al meglio le caratteristiche di entrambi.

A.2 Definizione di errore grossolano

Una efficace definizione intuitiva di errore grossolano si può ricavare dall'espressione inglese "outlier" (che letteralmente significa: che sta al di fuori), indicando con ciò un'osservazione che si stacca così nettamente dall'insieme delle altre, da indurre il sospetto che sia stata generata da un meccanismo differente. Invece se sono presenti più "outliers" raggruppati, l'ipotesi di un evento malaugurato è più difficile da sostenere; piuttosto si deve cercare di capire, perché è nato quel gruppo di osservazioni discordanti e come le possibili spiegazioni influenzano le precedenti conoscenze del fenomeno.

Occorre naturalmente distinguere fra analisi di serie di dati di produzione, con modello interpretativo già consolidato e collaudato, e serie di dati utilizzate nella fase di studio e costruzione di un modello, dove bisogna esaminare con cura anche ogni indicazione in contrasto con le ipotesi di partenza. In ogni caso, il problema degli "outliers" è un problema fisico (geometrico), cioè di comprensione del fenomeno, e non statistico quale effetto di certe osservazioni su uno stimatore o sulle stime ricavate da un insieme di dati. Parlare di un meccanismo diverso che genera "outliers" può significare non presenza di errori, ma proprio un comportamento diverso del fenomeno in condizioni diverse di osservazione. "Outlier" è un qualcosa che sta fuori rispetto ad un perimetro implicitamente definito dal modello (e, se questo è corretto, anche dalle osservazioni). Tuttavia accade spesso che molti modelli abbiano carattere empirico od approssimato e la presenza di valori contraddittori costituisca un punto di domanda: o si è in grado di mostrare che c'è un errore in quella particolare osservazione, oppure essa va spiegata in modo diverso, cioè con un modello diverso.

La definizione iniziale, per come è formulata, affida il riconoscimento dell'errore all'evidenza, ad esempio, alla visualizzazione dei dati: si tratta di un criterio certamente limitato. Volendo fissare, in termini quantitativi, l'appartenenza o meno di una osservazione all'insieme dei dati, si devono definire un parametro di centro

che rappresenti il nucleo dei dati ed una qualche misura di dispersione attorno ad esso. Questo consente di poter giudicare, in base ad una qualche ipotesi distribuzionale per le osservazioni, se quel dato è un "outlier" o un "inlier". Pertanto il problema dell'individuazione degli errori grossolani e del loro trattamento (eliminazione, sotto-pesatura, ecc.) dipende strettamente dalle ipotesi di modello e dal tipo di statistiche utilizzate.

Storicamente la prima risposta alla questione è stata (ed è tuttora anche se non è più l'unica) una reiezione soggettiva da parte del responsabile del trattamento dei dati. Completato un campione di osservazioni, prima di utilizzarle per ulteriori elaborazioni, esse vengono ispezionate, spesso visualizzandole graficamente, da una persona competente che, sulla base della propria esperienza e sensibilità, decide se vi sono dati da scartare, perché errati. Questo tipo di intervento soggettivo era l'unico praticabile prima che l'analisi statistica dei dati ed i risultati della teoria della stima fornissero un quadro più oggettivo in cui inserire la variabilità dei risultati di misura; esso può a buon diritto essere considerato il primo metodo robusto di stima.

Numerosi studi sono stati compiuti per valutare la concordanza dei risultati ottenuti con tali metodi: vari insiemi di dati, estratti da prefissate distribuzioni e contaminati in modo differente, sono stati affidati da esperti del settore, perché li esaminassero segnalando gli "outliers", confrontando poi le scelte fatte. I risultati non sono sempre univoci (in particolare in presenza di contaminazioni con "outliers" non evidenti): talvolta si sono ottenuti campioni puliti, altre volte è emerso un criterio di eliminazione non ben definito, basato sulla sensazione di aver di fronte un valore non convincente. Tuttavia uno screening preventivo, da parte di un esperto, resta una buona assicurazione sui guasti causati alle stime dagli "outliers". Attualmente la disponibilità di mezzi di calcolo sempre più potenti consente l'analisi di serie di dati assai estese e complesse. Giocoforza questa analisi finisce per avvenire senza un preventivo controllo a vista: per tanto si richiede alle metodologie statistiche adoperate, di svolgere anche questo ruolo di difesa dagli errori, in modo sempre più affidabile ed efficace.

Contrariamente a quanto si potrebbe pensare, lo sviluppo delle scienze statistiche, se da un lato ha introdotto una serie di criteri di oggettività (definendo le proprietà degli stimatori e gli indici per confrontarne le prestazioni), dall'altro mantiene sostanzialmente inalterata la soggettività e le lacune nell'affrontare il problema degli "outliers". Esso viene solo spostato dalla decisione sul dato da escludere alla scelta del metodo di stima e di reiezione da impiegare, e non solo perché esistono più metodi per interpretare i dati, ma anche e soprattutto perché, fino alla seconda metà degli anni sessanta, l'analisi della sensibilità agli errori grossolani dei diversi metodi di stima, fino ad allora elaborati, è stata scarsa ed insufficiente. La questione degli "outliers" era relegata ad un pretrattamento (empirico) delle osservazioni e sembrava non dover essere integrata a pieno titolo nell'analisi statistica e tra le proprietà degli stimatori.

In effetti, le proprietà di ottimalità dei metodi di stima classici (massima verosimiglianza, minima varianza, minimi quadrati) sono sull'ipotesi che il campione estratto appartenga ad una certa classe di distribuzioni di probabilità: se invece il campione è contaminato, il comportamento degli stimatori si rivela spesso assai inaffidabile, portando a stime fortemente deviate. Il problema chiave, per questi stimatori, è che essi non operano discriminazioni fra i dati e costruiscono una stima che, per quanto si introducano pesi, è strutturalmente indifesa contro gli "outliers". Si consideri la media aritmetica: poiché essa costituisce il baricentro delle osservazioni, risulta fortemente deviata da valori anomali e non coglie dove sta il grosso dei dati. Questo implica che anche il grado di dispersione, calcolato rispetto a quel centro, sarà falsato e ogni meccanismo di decisione, basato sulla distanza standardizzata rispetto a quel centro, sarà

costituzionalmente debole.

E' questa una delle critiche fondamentali al metodo dei minimi quadrati, avanzate dai proponenti di nuove procedure di stima, dette robuste, la cui motivazione è proprio la capacità di fornire stime, magari meno efficienti, ma certo assai meno deviate, quando le ipotesi del modello sono in qualche modo violate nel campione. Un quadro delle caratteristiche di queste procedure verrà esposto più avanti; i cenni fatti avevano lo scopo di mostrare come, anche da un punto di vista strettamente statistico, il problema degli "outliers" sia ancora aperto e soprattutto che la nozione stessa di "outliers" è in qualche modo ambigua, in quanto dipende, in modo essenziale, dal tipo di ipotesi e scelte che si fanno nel trattamento dei dati.

A.3 Generazione degli "outliers"

Da un punto di vista statistico due meccanismi base portano alla formazione di campioni in cui si sospettano presenti degli "outliers"; può essere di una certa importanza, per il successivo trattamento dei dati, riconoscere quale dei due agisce in un certo insieme di osservazioni.

- I dati provengono da qualche distribuzione a code piene: in tal caso, nessuna delle osservazioni può essere definita erronea in senso stretto, cioè generata da un meccanismo diverso rispetto alle altre, poiché in queste distribuzioni la probabilità di eventi lontani dal centro non è trascurabile. Di conseguenza tutti i dati forniscono informazione e non si deve porre la questione della reiezione di tali osservazioni, bensì quello di un loro adeguato trattamento. Un tentativo di classificare le caratteristiche delle distribuzioni, in base alla loro sensibilità alla presenza di "outliers", è stato fatto distinguendo fra le famiglie di distribuzioni "outlier-prone" e "outlier-resistenti". Come intuitivo, le prime hanno code che vanno a zero lentamente, le seconde rapidamente. La definizione introdotta prende in considerazione la distribuzione della differenza e del quoziente fra gli ultimi due elementi di un campione ordinato (il massimo ed il valore precedente) e ne considera il comportamento al crescere della numerosità del campione.
- I dati provengono da due distribuzioni differenti di cui una, detta distribuzione base, genera le osservazioni corrette mentre l'altra, la distribuzione contaminante, fornisce dati provenienti da un fenomeno estraneo o comunque diverso rispetto al modello base. Anche in assenza di una distanza significativa fra i centri delle due distribuzioni, se la seconda è nettamente più a code piene della prima, si ha una tendenza all'apparire di "outliers". Ad esempio, un meccanismo del genere può essere invocato per descrivere la presenza di "blunders", cioè di valori del tutto estranei alle osservabili. Un altro caso è quello in cui è il modello deterministico scelto ad essere sbagliato: in situazioni del genere, parecchie osservazioni vengono viste dal modello come apparenti "outliers". Un'ultima categoria è quella in cui vi sono effetti, transitori o secondari nel contesto generale del fenomeno, che agiscono solo su una parte dei dati, alterando i valori rispetto a quanto previsto dal modello teorico. L'esistenza di una contaminazione delle osservazioni può essere affrontata in due condizioni differenti; conoscendo in partenza, su un campione di una data numerosità, il numero di "outliers" presenti, oppure no. In questo secondo caso, si può asserire che esiste una probabilità che, ad ogni estrazione, il dato appartenga alla distribuzione contaminante e, per converso, che sia il complemento ad uno la probabilità di estrazione da quella base. In tal modo, il numero di "outliers" diventa una variabile casuale di tipo binomiale e si può calcolare la probabilità di accadimento di "outliers" su un dato numero di estrazioni. Naturalmente il

valore di probabilità di tale accadimento di “outliers” è incognito, ma in generale è sensato assumerlo prossimo a zero. Pertanto il problema diventa, nei modelli di tipo parametrico, passare alla stima di due parametri, uno per ciascuna delle due distribuzioni di probabilità (ovvero del loro rispettivo peso) nel modello misto prescelto, ed eventualmente studiare la variabilità dei risultati di funzione della probabilità di accadimento di “outliers”.

A.4 Considerazione sulle ipotesi di modello

Il modello funzionale ed il modello stocastico sono introdotti e giustificati sulla base di una serie di affermazioni di carattere generale, date per valide. Tuttavia è necessario analizzare, più a fondo, il significato dei modelli ed il procedimento con cui li si formula, per capire meglio in quale modo i risultati dipendono dalle ipotesi fatte e soprattutto come vengono alterati, se una qualche ipotesi non è soddisfatta.

Il modello funzionale esprime relazioni matematiche tra grandezze teoriche, le osservabili. Indipendentemente dalla forma con cui tali relazioni sono espresse, cioè come equazioni di condizione pure o attraverso l'introduzione di parametri, se la formulazione analitica non è contraddittoria, il modello risulta in sé consistente (da un punto di vista numerico, nella scrittura delle equazioni si deve evitare la possibilità di introdurre malcondizionamenti nella matrice normale del sistema, per non avere stime con varianza elevata). Tuttavia rimane aperto il problema della validità delle relazioni, cioè della reale capacità della teoria di spiegare il fenomeno tenendo conto di tutti i fattori che lo influenzano, almeno entro la soglia della sensibilità impiegata nell'acquisizione dei dati. In generale, quando si vuole costruire un modello matematico per spiegare un certo fenomeno fisico (geometrico), si parte dalla formulazione di un modello semplice, dove le modalità operative o le incertezze dovute all'acquisizione dei dati non vengono introdotte. In questa fase il problema è di includere tutte le grandezze fisiche, significativamente in gioco, o di scegliere una geometria, sufficientemente approssimata, per descrivere il problema.

Il secondo passo è più strettamente legato alle modalità operative ed alle incertezze insite nell'acquisizione dei dati: in pratica, si tratta di introdurre parametri che tengano conto di effetti sistematici o modifiche alla geometria ideale in cui il modello è costruito. L'inserimento di parametri aggiuntivi, per tener conto di effetti sistematici, è probabilmente la fase più delicata di affinamento del modello. Infatti spesso, la formulazione matematica di tali effetti è difficile da trovare ed è affidata a funzioni semplici (polinomi di basso grado) o ricavate empiricamente (con relazioni di dubbia validità generale). Infine tanto più l'effetto da modellare è di modesta entità, tanto meno significativi risultano i valori stimati ed è dubbia l'opportunità della loro presenza nel modello.

Un ultimo problema è quello di verificare il comportamento delle stime quando, come nella maggior parte dei casi, sia necessario linearizzare le relazioni. In tal caso, le stime, oltre a perdere le proprietà di ottimalità, in generale, dipendono dall'insieme di valori approssimati di cui si dispone.

Il modello stocastico è anch'esso soggetto ad incertezza. Anzitutto è quasi sempre necessario supporre l'indipendenza fra le osservazioni (anche se talora le metodologie operative sono tali da implicare correlazioni), perché sarebbe assai difficile valutarle a priori. Inoltre le ipotesi sulla varianza della singola osservazione sono legate quasi esclusivamente alla sensibilità impiegata nell'acquisizione dei dati, non essendo possibile volta a volta quantificare l'effetto sulla varianza delle condizioni operative di quella particolare osservazione.

Il punto critico del procedimento sta comunque nella identificazione fra osservabili ed osservazioni: si vuole cioè adattare un modello teorico, che una volta formulato è rigidamente definito, ad un insieme di dati che sono il risultato di operazioni sperimentali. Il valore numerico dell'osservazione risente di una serie di fattori, detti errori di osservazione, di cui si conosce solo in parte il numero e mai l'entità dell'effetto (salvo per un comportamento tendenziale nell'ipotesi che la perturbazione al valore teorico dell'osservazione sia di tipo accidentale, cosa che comunque non consente di dire nulla sulla singola osservazione). La definizione del modello stocastico cerca proprio di tener conto di tale natura dei valori sperimentali. Tuttavia le circostanze della singola operazione, piuttosto che di un certo insieme di osservazioni, possono alterare le osservazioni in misura assai superiore a quella prevista dal modello, e portare ad un valore di sigma zero maggiore di quanto atteso, dato lo schema di misura e le incertezze dovute all'acquisizione dei dati.

In generale, si parla di effetti sistematici quando si rileva per tutte le osservazioni, o comunque in un certo insieme di esse, uno scostamento rispetto al modello teorico che ha un comportamento analiticamente determinato. In tal caso, è la varietà dei valori ammissibili che dovrebbe essere riformulata, almeno parzialmente, introducendo nuovi parametri, perché è evidente che, in questo caso, si ha un spostamento delle medie teoriche delle osservabili.

In presenza di errori grossolani, è possibile scegliere di ascrivere all'una od all'altra parte del modello la discrepanza. In linea di principio, si può eliminare l'errore inserendo un parametro aggiuntivo nella equazione corrispondente, il che significa dire che quella osservazione è stata fatta su una osservabile diversa da quella ipotizzata, spostata nella media di un valore pari a quello del parametro. D'altra parte, lo stesso errore può essere interpretato come dovuto ad una sottovalutazione della varianza, cioè ad una estrazione da una variabile casuale con la stessa media, ma assai più dispersa del previsto. In entrambi i casi comunque, le stime risultano distorte (in quanto vincolate ad una varietà lineare che non è quella delle osservabili effettive), oppure più disperse del previsto (perché la direzione di proiezione sulla varietà non è quella corretta).

Da un punto di vista operativo, si richiede una metodologia che consenta di valutare la correttezza del modello, in relazione all'attuale vettore di osservazioni, cioè di segnalare la presenza ed identificare effetti sistematici ed errori grossolani. Da un punto di vista progettuale, è utile poter sapere quanto sia affidabile un certo schema di misure, cioè quale possibilità offra di identificare eventuali errori grossolani, attraverso la misura dell'affidabilità (controllabilità locale) delle osservazioni.

Nel primo caso, vanno costruiti test statistici che consentano di rilevare se, nel complesso delle osservazioni, esistano errori grossolani ed, in subordine, di riconoscere quali siano le osservazioni sospette. Nel secondo caso, è importante poter valutare (a priori) l'entità del minimo errore grossolano evidenziabile (ovvero l'affidabilità interna dello schema di misura) e, in presenza di tale errore, la distorsione delle stime dei parametri (ovvero l'affidabilità esterna).

A 5 Verifica di ipotesi parametriche

La stima con il metodo dei minimi quadrati non richiede ipotesi sulla funzione di distribuzione delle osservazioni. D'altra parte la natura stocastica delle osservazioni impone una valutazione dei risultati che deve essere in definitiva di tipo probabilistico. Infatti è possibile proporre come criterio per valutare se uno scarto od un parametro stimati sono significativi (cioè sensibilmente diversi da zero), il rapporto fra la stima in esame ed il suo sqm, anche se è chiaro che, in ultima analisi, la scelta del valore soglia, per la decisione,

finisce per essere empirica ed arbitraria. Invece data una funzione di distribuzione delle osservabili, si possono utilizzare i metodi dell'inferenza statistica, in particolare di quella parametrica, per verificare ipotesi sui risultati.

In sintesi, si abbia un'ipotesi fondamentale (o nulla) per il valore di un parametro o di uno scarto che individua, all'interno di una famiglia, una particolare distribuzione, e si abbia un valore campionario dello stesso: scopo del confronto di ipotesi è indicare se la differenza fra i valori è significativa in termini probabilistici. Fissato un livello di significatività del test statistico e stabilita, se necessario, un'ipotesi alternativa, l'ipotesi fondamentale verrà accolta, se il valore campionario considerato rientra nella regione di accettazione determinata, intorno all'ipotesi fondamentale, dalla forma della distribuzione e dal valore del livello di significatività. In presenza di ipotesi alternativa, si ricava anche il cosiddetto rischio del consumatore (contrapposto al livello di significatività, rischio del produttore) ed il suo complemento all'unità è detto potenza del test statistico. Il valore del livello di significatività è la probabilità che si ha, eseguendo il test statistico, di respingere un'ipotesi fondamentale vera (errore di I tipo); il valore della potenza del test statistico è la probabilità di accettare un'ipotesi fondamentale falsa (errore di II tipo). Fissato il livello di significatività, resta individuato l'intervallo di confidenza del valore campionario, cioè l'arco di valori compatibili, a quel livello di significatività, con il particolare valore campionario estratto; la potenza del test statistico indica invece la capacità di distinguere fra ipotesi alternative che meno differiscono, tanto maggiore è la potenza del test statistico (avendo a disposizione più tipi di test statistici, è preferibile usare quelli con curva di potenza più rapida)

L'ipotesi comunemente accettata per le osservabili, è che esse provengano da una distribuzione normale multi – dimensionale, le cui componenti siano indipendenti. Questa scelta ha un supporto empirico in una lunga serie di risultati sperimentali in ambiti di misura assai differenti ed uno fondamentale, di tipo teorico, nel teorema di Gauss (o limite centrale della statistica). Le proprietà delle distribuzioni normali consentono un'agevole determinazione, almeno nell'ambito dei modelli lineari, delle distribuzioni campionarie degli stimatori che interessano, così che ad essi può essere estesa l'inferenza statistica. Per i modelli non – lineari esiste una teoria piuttosto complessa; tuttavia le conclusioni trovate in ambito lineare saranno valide, in forma approssimata, nei casi in cui la linearizzazione delle equazioni di osservazione non si stacca molto dal modello nel dominio di maggior densità delle osservabili.

A.6 Definizione dell'ipotesi fondamentale

Schematicamente si possono raggruppare in tre categorie le principali cause d'errore rilevabili nel trattamento dei dati:

- ❑ errori nel modello funzionale;
- ❑ errori nel modello stocastico;
- ❑ problemi numerici.

La distinzione fra i tre gruppi non è sempre netta, poiché talvolta la medesima causa può essere attribuita a più categorie. Per quanto riguarda il modello funzionale, si hanno:

- ❑ errori nella linearizzazione o nel calcolo dei termini noti (questo tipo di errore si ha ovviamente solo nella

fase di messa a punto del modello e non in quella applicativa);

- ❑ valori approssimati dei parametri non sufficientemente precisi (è noto che, se la linearizzazione è fatta attorno ad un punto lontano dalla soluzione, i metodi iterativi possono non convergere o convergere ad altre soluzioni);
- ❑ valori approssimati incompatibili con il modello parametrico prescelto (le equazioni di osservazione sono linearizzate entro certi sistemi di riferimento);
- ❑ problemi in cui la parametrizzazione introdotta ha una deficienza di rango, numero o tipo di vincoli insufficiente o inadeguato (ad esempio, quando occorre fissare un sistema di riferimento);
- ❑ inadeguatezza del modello a descrivere il fenomeno con una buona approssimazione, cioè necessità di introdurre altri parametri per spiegare le osservazioni (in questi casi, si hanno effetti di tipo sistematico non modellati che confluiscono negli scarti, compromettendo la precisione);
- ❑ parametri mal stimabili con l'insieme di dati a disposizione: casi di questo tipo sono dovuti ad una geometria delle osservazioni non sufficientemente collaudata in fase di simulazione, per cui il modello è debole localmente (ad esempio, nelle strutture reticolari, si hanno casi in cui le osservazioni sono insufficienti a determinare un parametro, ma il procedimento numerico arriva egualmente al calcolo di una soluzione che tuttavia, essendo localmente indeterminata, ha varianza molto elevata; in altri casi, ad esempio nelle interpolazioni, si possono avere coefficienti con termini molto vicini a zero: si ha pertanto una riga ed una colonna della matrice normale quasi nulle, cioè un determinante molto piccolo che crea instabilità nella sua inversione);
- ❑ parametri non distinguibili con l'insieme di dati a disposizione (ad esempio, quando il comportamento di due o più parametri si differenzia poco nell'ambito delle osservazioni: questo fa sì che due righe e due colonne della matrice normale risultino circa uguali e, come nel caso precedente, il determinante sia molto piccolo);
- ❑ errori grossolani nelle osservazioni (come si è già detto, la loro attribuzione al modello deterministico impone che nella osservazione contenente l'errore si introduca un opportuno parametro di servizio).

Per quanto riguarda il modello stocastico, si hanno:

- ❑ cattiva scelta dei pesi (ad esempio, dovuta ad una valutazione o ad una progettazione inadeguata dei rapporti di precisione, fra gruppi di osservazioni o all'interno di uno stesso gruppo; nel primo caso, si fa prevalere un tipo di misure sulle altre, nel secondo caso, si può avere una debolezza diffusa dello schema di misura);
- ❑ presenza di correlazioni insospettite fra le osservazioni (ben difficilmente si può disporre di una valutazione empirica di tali correlazioni, pertanto si accantona il problema, presupponendo la indipendenza; in generale, queste situazioni si verificano in presenza sistematici non modellati);
- ❑ errori grossolani nelle osservazioni (dal punto di vista del modello stocastico, si tratta di attribuire un peso assai più basso all'osservazione corrispondente, per ridurre l'effetto dello scarto sulla stima di sigma zero).

Per quanto riguarda i problemi di tipo numerico, non è facile dare un elenco preciso di cause ed effetti. Come osservazione preliminare, valida per ogni tipo di algoritmo numerico, è necessario valutare, in

relazione all'approssimazione richiesta per le stime attese, l'effetto degli errori di arrotondamento causati dalle diverse fasi del calcolo. Può essere perciò indispensabile, accanto alla considerazione del numero di cifre significative garantite, utilizzare tecniche di soluzione, forse meno efficienti dal punto di vista dei tempi di calcolo e della occupazione di memoria, ma che introducono errori minori. In ogni caso, il buon condizionamento della matrice normale del sistema è uno dei punti chiave per ottenere la stabilità della sua inversione, ove necessaria. A questo scopo, è utile calcolare, nel corso dell'elaborazione, il cosiddetto numero di condizione.

A.7. Le stime in norma-1

Il criterio di stima dei minimi quadrati, introdotto da Gauss alla fine del '700, è probabilmente il più usato e studiato in statistica. Tuttavia storicamente però non è il primo a nascere: intorno al 1750, Boscovitch propose, come stimatore del nucleo delle osservazioni, il vettore \bar{x} che rende minima la somma dei moduli degli scarti fra osservazioni e modello.

Si tratta di un principio basato sulla cosiddetta norma-1 (somma dei moduli delle componenti), in alternativa alla usuale norma-2 (o norma euclidea, radice quadrata della somma dei quadrati delle componenti). Fra le sue proprietà fondamentali, il metodo ha quella di interpolare in modo esatto tante osservazioni quanti sono i parametri da stimare, cosicché altrettanti residui risultano sempre identicamente nulli. Una serie di difficoltà, legate ai metodi di calcolo della soluzione nell'ambito dei modelli lineari, lo pose in secondo piano rispetto ai minimi quadrati. Tuttavia fin da principio, fu notato che, in presenza di contaminazioni entro le osservazioni, il metodo era assai più efficace di quello proposto da Gauss.

Probabilmente due furono le principali ragioni del successo incontrato dei minimi quadrati, rispetto a questo e ad altri metodi di stima: da una parte gli studi delle distribuzioni sperimentali degli errori accidentali di misura, dall'altra lo sviluppo di una teoria della stima e la formulazione del criterio di massima verosimiglianza. Pertanto i modelli parametrici normali risultavano giustificati empiricamente e ottimali teoricamente.

Entrambe queste motivazioni, a partire dagli anni '60 (fondamentale è uno studio di Tukey, che eseguì numerose prove su campioni provenienti da distribuzioni contaminate), furono sottoposte a critica, quando l'attenzione si concentrò sul problema degli "outliers". Nel tentativo di introdurre nuovi criteri di valutazione delle proprietà degli stimatori, in presenza di deviazioni dal modello teorico assunto, Huber e Hampel che possono essere considerati i padri fondatori di queste nuove procedure, verso la metà degli anni '60, proposero la robustezza quale ulteriore e fondamentale requisito per gli stimatori.

A.8 Obiezioni all'uso di modelli parametrici esatti

La critica fondamentale, rivolta ai metodi classici di stima, è che essi derivano procedure ottimali sotto tutta una serie di ipotesi che, per lo più, non è possibile verificare nel singolo caso e che vengono date per valide comunque. Poiché tali teorie non forniscono informazioni sul comportamento delle stime quando, come è praticamente sempre vero, la realtà mostra deviazioni rispetto alle assunzioni fatte, non c'è garanzia di validità sui risultati.

Una tacita speranza nel trascurare le deviazioni dai modelli ideali era che esse non avessero importanza: che le procedure statistiche ottimali sotto l'ipotesi di modello

esatto fossero approssimativamente ottimali sotto un modello approssimato. Sfortunatamente si è visto che tale speranza era spesso drasticamente mal riposta; persino leggere deviazioni hanno spesso effetti molto più estesi di quanto fosse previsto dalla maggior parte degli statistici (Tukey, 1960).

Si possono distinguere quattro categorie principali di deviazione dal modello parametrico:

- ❑ presenza dei grandi errori grossolani;
 - ❑ modello concepito come una approssimazione;
 - ❑ effetti dovuti agli arrotondamenti e al raggruppamento dei dati;
 - ❑ indipendenza delle osservazioni (o più specificamente, una certa struttura di correlazione) solo parzialmente vera.
-
- ❑ La prima circostanza deriva dal fatto che, a formare l'insieme di valori sottoposti allo "screening" del modello, non concorrono solo gli errori di misura in senso stretto. Se per questi, a prescindere dallo specifico tipo di distribuzione (se normale o moderatamente leptocurtica), è trascurabile la probabilità di un errore grossolano, altri fattori presenti (errori di trascrizione, ecc..) hanno invece una dispersione estremamente superiore. La presenza di errori grossolani è certamente la più dannosa fra i tipi di deviazione sopra esposti: è sufficiente non identificare uno solo di questi, per rovinare completamente l'analisi di dati. D'altra parte, poiché è del tutto comune ritrovarne percentuali di qualche unità, entro masse di dati non preventivamente controllate, ciò è particolarmente pericoloso, quando fra i dati ed i risultati avviene solo un filtro operato automaticamente, se il metodo di stima usato non è robusto.
-
- ❑ La seconda categoria di deviazioni dal modello è la più complessa da analizzare e, in parte, è collegata con la prima, essendo la distinzione fra grandi errori grossolani ed "outliers" moderati non sempre ben definiti. La approssimazione del modello può essere di due tipi, a livello del modello funzionale ed a livello del modello stocastico. Molti modelli hanno un carattere empirico; spesso quelli di tipo lineare, sono in realtà approssimazioni (linearizzazioni), rispetto al comportamento teorico. Esistono fenomeni dove effetti transitori provocano variazioni rispetto alla fase di regime; infine il modello può essere semplicemente errato perché in fase di elaborazione. Pertanto non è facile dire se un "outlier" apparente lo sia effettivamente; gli scostamenti fra osservazioni e valori calcolati in un modello errato tendono a diventare "outlier". Anche le caratteristiche distribuzionali degli errori sono messe in questione da accurati studi, effettuati su campioni di elevata qualità, ovvero dopo preventiva ripulitura dagli evidenti "outlier": in moltissimi casi la curva normale non è la curva che meglio si adatta all'istogramma dei dati, risultando le distribuzioni sperimentali moderatamente leptocurtiche. E' interessante, sotto questo punto di vista, richiamare le affermazioni dello stesso Gauss sulla genesi dei minimi quadrati:

L'autore del presente trattato, che nel 1797 per primo studiò questo problema alla luce della teoria della probabilità, si

accorse ben presto che era impossibile determinare il valore più probabile della quantità incognita, senza che fosse nota la distribuzione degli errori. Tuttavia poiché questa non lo è, non c'era altra possibilità che assumere tale funzione per ipotesi. All'autore, sembrò la cosa più naturale prendere l'approccio opposto, cercando quella funzione che deve essere presa come base affinché, nel più semplice dei casi, si ottenga una regola che è generalmente accettata come valida, cioè che la media aritmetica di diverse osservazioni di pari precisione, per una specifica grandezza, dovesse essere considerata il valore più probabile (Gauss,1821).

Poiché tale criterio conduce proprio ad una distribuzione di tipo normale, è evidente che Gauss prefissò in realtà lo stimatore, non il modello probabilistico. Certamente un supporto consistente all'ipotesi di normalità è dato dal teorema di Gauss (o limite centrale della statistica), ma in ogni caso esso resta un teorema limite e può al più suggerire un comportamento approssimativamente normale, non certo garantirlo.

- I valori osservati hanno, per ragioni di sensibilità impiegata nell'acquisizione dei dati, un numero finito di cifre significative, cioè sono intrinsecamente discretizzati. Spesso inoltre essi sono arrotondati, raggruppati e classificati. La presenza di imprecisioni localizzate entro la scala di misura, può provocare piccoli effetti sistematici, in generale trascurabili, ma che hanno influenza in ambiti specifici. Lo studio di queste proprietà di granularità delle stime, cioè della sensibilità a piccole variazioni dei dati, può essere fatto utilizzando un indice introdotto dalla teoria della robustezza, detto "local shift sensitivity". Esso mostra come in questi casi, stimatori robusti rispetto agli "outliers", quali la mediana, siano decisamente sensibili, mentre la media aritmetica, notevolmente debole in presenza di "outliers", si comporti assai bene.

- Per quanto riguarda infine la presenza di correlazioni insospettite, il campo di indagine sugli effetti prodotti sulle stime è ancora molto aperto, essendo pochi e specifici gli studi in materia; pertanto ancora meno si conosce su procedure robuste rispetto a violazioni dell'indipendenza. In ogni caso, si tenga presente che la presenza di correlazioni fra le osservazioni costituisce, in generale, un notevole deterioramento della precisione della stima dei parametri del problema in esame.

A.9 La teoria della robustezza

Molte sono definizioni relative alla robustezza ed alle statistiche robuste. Citando Galileo Galilei (nel Dialogo sui massimi sistemi) ed i due principali autori moderni ai quali è dovuto lo sviluppo di questa teoria, si può dire che:

Per poter ritrarre (ottenere) delle osservazioni più notizie (informazioni) che sia possibile, conviene emendare

(correggere) i loro errori; cosa conveniente è che noi affidiamo le minori e più vicine emenda (correzioni) dove si possa (Galilei,1632).

Robustezza significa insensibilità rispetto a piccole deviazioni dalle assunzioni (Huber, 1977).

Le statistiche robuste, viste come collezione di teorie connesse, sono le statistiche dei modelli parametrici approssimati (Hampel, 1986).

Per capire, almeno intuitivamente, cosa si intende per piccole deviazioni o modelli approssimati, conviene evidenziare il diverso comportamento delle stime parametriche e non. Dato uno spazio campionario e lo spazio infinito – dimensionale di tutte le possibili distribuzioni di probabilità definibili al suo interno, le statistiche non-parametriche riducono la scelta delle possibili distribuzioni ad un sottospazio ancora piuttosto esteso; un modello parametrico invece definisce un sottospazio caratterizzato da poche dimensioni (i gradi di libertà dei parametri da stimare). Le statistiche robuste infine rendono disponibile un intorno infinito – dimensionale di un particolare modello parametrico, cioè lo spazio delle distribuzioni di probabilità vicine ad esso. Infatti non si tratta di mettere in discussione la validità dei modelli parametrici, bensì di cercare metodi adatti a lavorare quando la reale distribuzioni sottostante il campione sia poco lontana da quella ipotizzata. I principali scopi delle statistiche robuste possono perciò essere riassunti così:

- ❑ descrivere la struttura che meglio al grosso dei dati;
- ❑ identificare chi sta al di fuori di tale nucleo (“outliers”, sotto-strutture);
- ❑ in presenza di strutture di dati non bilanciate rispetto al modello, identificare punti con elevata influenza sulle stime (punti di leva).

- ❑ Si tratta di assumere un modello parametrico tenendo esplicitamente in conto il fatto che esso può essere distorto (o se si vuole che una parte dei dati può non appartenere al modello. Questo segna un orientamento diverso rispetto alle stime minimi quadrati, dove l’analisi di congruenza delle osservazioni è a posteriori (e pertanto già distorta) e ciascun dato ha un peso uguale nel costruire il valore della stima. Il problema è stabilire cosa significa, di volta in volta, il grosso dei dati. La massima percentuale di dati estranei, ovvero di minoranza, ammessa da uno stimatore, prima che essa provochi sulla stima effetti superiori ad una certa soglia, conduce a definire il cosiddetto punto di rottura (in inglese: breakdown point). La decisione su quale modello scegliere deve essere presa con una verifica di ipotesi, ad esempio, per confronto con un modello esteso (come per i modelli lineari), nel qual caso si tratta di ipotesi di nullità per certi parametri.
- ❑ I residui, rispetto ad un centro robusto, evidenziano gli “outliers” meglio di quelli dei minimi quadrati, anche se si usano tecniche inferenziali quali il massimo residuo standardizzato. Successivamente bisognerebbe cercare di capire come sono nati gli errori (e possibilmente correggerli) anziché procedere alla eliminazione. Sotto questo punto di vista, in molti metodi di reiezione sembra prevalere il timore di

rovinare le statistiche, più che la ricerca di una reale giustificazione per eliminare una osservazione. L'usuale argomentazione che valori sulle code hanno probabilità quasi nulle e quindi nascono da altri fenomeni certo non è infondata, ma sembra strettamente legata alla debolezza di certe statistiche, esistendo metodi in cui gli "outliers" hanno poca influenza sui parametri di locazione (in questo caso, la funzione d'influenza degli stimatori è limitata o addirittura decrescente). Infine bisogna richiamare il fatto che spesso esiste una regione di dubbio, nella quale sono presenti dati che non sono evidenti outlier e neppure valori buoni: alcune procedure robuste sono pensate proprio per tener conto di questa situazione.

- Non tutti i punti di osservazione del fenomeno hanno la stessa importanza nella determinazione delle stime; questo vuol dire che un errore in quelle osservazioni può alterare in modo rilevante il risultato. Nella regressione lineare semplice, ad esempio, se un valore è misurato molto lontano dagli altri, bisogna anzitutto chiedersi se usarlo o meno, perché la sua sola presenza determina la posizione della retta interpolatrice molto più di un qualsiasi altro punto. Poiché lo stesso fenomeno si ha anche a più dimensioni, dove un controllo qualitativo è più difficile, parecchie procedure robuste evidenziano questi cosiddetti punti di leva.

A.10 Una strategia per l'integrazione fra soluzioni L_1 (o altre procedure robuste) e L_2

Dato che è possibile calcolare la soluzione di un problema sovradeterminato anche con procedure robuste, cioè poco deviate da errori grossolani, è spontaneo chiedersi se, non potendosi ragionevolmente asserire che tutte le osservazioni scartate siano tout-court errate, esista un modo per recuperare la maggior parte dell'informazione (apparentemente) trascurata dalla soluzione. Altrettanto importante (e anzi preliminare) è la questione su quale giudizio dare della precisione dei risultati.

In questo caso, un serio ostacolo è dato dal fatto che non si conoscono le distribuzioni campionarie delle stime ricavate con procedure robuste. Infatti qualunque sia la distribuzione ipotizzata per le osservabili, il fatto che la soluzione non possa essere espressa da una relazione chiusa (anche non – lineare) funzione delle osservazioni impedisce l'applicazione della legge di propagazione della covarianza ed il calcolo della matrice di covarianza, così come avviene per i minimi quadrati. Poiché non si può rinunciare ad una valutazione dell'incertezza, si è costretti a tornare, con qualche compromesso, ai minimi quadrati. Pertanto si ripropone il problema di recuperare le osservazioni accettabili, per non penalizzare la ridondanza e, di conseguenza, la qualità delle stime effettuate.

Per compiere questo percorso a ritroso, una prima idea è quella di prefissare un tetto al numero di "outliers" presenti. Stabilita la percentuale di "outliers" che, per quel tipo di problemi e per il pretrattamento eventuale delle misure, si ritiene non possa essere ragionevolmente superato, si scartano le osservazioni con i residui maggiori e si processano, a minimi quadrati, le restanti. La logica di questa idea (pagare un premio contro gli errori) per una assicurazione è la stessa vista all'inizio, a proposito delle distribuzioni contaminate, quando la percentuale di errore è decisa (a priori).

Si può obiettare che, dopo una procedura robusta, una rielezione ad occhi chiusi non è giustificata. Allora un altro criterio è quello di guardare gli scarti della soluzione L_1 (o di altre procedure robuste) e decidere, con un criterio empirico, quali osservazioni scartare. Poiché le procedure robuste individuano il nocciolo delle osservazioni concordanti, è giustificato pensare che gli scarti più grandi siano quelli dovuti agli "outliers", mentre per le osservazioni con solo errori accidentali gli scarti siano relativamente piccoli. Una prima idea

può essere quella di costruire l'istogramma degli scarti, suddividendoli in classi, e vedere se questo presenta un punto di taglio fra "outliers" e osservazioni pienamente accettabili. Individuato tale punto, si scartano i valori più elevati.

In alternativa, si può fissare una unità di misura (ed un suo multiplo, la tolleranza) per decidere in relazione a cosa uno scarto è grande o piccolo. Si noti che la stima di sigma zero non è adatta, perché essendo una media dei moduli degli scarti non nulli, non è robusta. Pertanto conviene assumere, quale misura di dispersione, la mediana dei moduli degli scarti; tuttavia la scelta della tolleranza resta arbitraria, perché non può essere appoggiata ad alcun criterio probabilistico. Terminata questa eliminazione che, in ogni caso, riguarda una piccola frazione della ridondanza, si può eseguire la compensazione minimi quadrati delle osservazioni rimaste (e valutare le precisioni nella determinazione dei parametri) assumendo naturalmente come valore approssimato delle incognite i risultati della procedura robusta adottata.

E' possibile fermarsi a questo punto, poiché ragionevolmente la ridondanza non è stata troppo penalizzata e la precisione non ne ha risentito, in maniera eccessiva. Tuttavia se il filtro della procedura robusta adottata è stato efficace, si è in condizioni ottimali per adoperare i test specifici per i problemi ai minimi quadrati, perché non dovrebbero esserci effetti di mascheramento. Allora si può pensare di reintrodurre sequenzialmente almeno la parte migliore delle osservazioni con i residui più elevati, per cercare di migliorare la ridondanza, anche solo localmente, dove la debolezza dello schema, oppure il suo indebolimento in seguito alla reiezione, lo renda opportuno.

APPENDICE B – LA TEORIA DELLA ROBUSTEZZA¹⁴

B.1 L'approccio di Huber

Il primo a formalizzare un approccio alle statistiche dei modelli parametrici approssimati è stato Huber che ha introdotto una classe di stimatori, detti stimatori di tipo M (generalizzazione di quelli di massima verosimiglianza, detti stimatori di tipo L), e derivato la loro consistenza ed asintotica normalità.

Supposto, al solito, di avere n osservazioni x_i indipendenti, identicamente distribuite, con densità di probabilità $F(x_i - \theta)$, invece di massimizzare l'usuale funzione di verosimiglianza:

$-\sum_i \ln F(x_i - T) = \min$, ovvero di annullarne il differenziale totale $\sum_i F^T / F(x_i - T) = 0$, essendo T la

stima di ψ , Huber massimizza, semplicemente, la funzione obiettivo: $\sum_i \phi(x_i - T) = \min$, ovvero annulla

funzione d'influenza: $\sum_i \psi(x_i - T) = 0$, senza assumere alcunché su ϕ e ψ come funzioni di F , anzi

senza, neppure, richiedere che la funzione F sia la densità di probabilità di x .

Per tener conto della contaminazione, Huber elabora il cosiddetto "gross error model": invece di fissare $G(x - \theta)$ con G noto (es. la distribuzione normale), egli assume che una frazione nota ε ($0 \leq \varepsilon \leq 1$) dei dati provenga da una funzione di distribuzione di probabilità incognita $H(x - \theta)$. In definitiva, l'effettiva densità di probabilità F è data dal modello: $F(x - \theta) = (1 - \varepsilon)G(x - \theta) + \varepsilon H(x - \theta)$.

¹⁴ Questa appendice è liberamente tratta da: G.Forlani (1990): Metodi robusti di stima in geodesia e fotogrammetria. In: *Ricerche di Geodesia Topografia e Fotogrammetria*, n.8, CLUP, Milano.

Si tratta, a ben vedere, di un vero e proprio intorno del modello parametrico di base G , in quanto la presenza di H incognita (contrariamente ad altri modelli di contaminazione in cui, anche, la funzione di distribuzione di probabilità inquinante è data) definisce una regione dello spazio delle funzioni di distribuzione di probabilità, attorno a $G(x - \theta)$, la cui ampiezza è funzione di ε . Il passo successivo è l'ottimizzazione del comportamento di uno stimatore di tipo M , in questo intorno del modello, misurandone la stabilità attraverso la varianza asintotica $\sigma^2(\psi, F)$. Huber mostra che, sotto certe condizioni di regolarità per la funzione di distribuzione di probabilità F , tale limite può essere calcolato, come varianza dello stimatore stesso, e che, se si assume l'esistenza di un legame fra la funzione d'influenza e l'effettiva densità di probabilità, la varianza asintotica risulta:

$$\sigma^2(\psi, F) = \frac{1}{I(F)}$$

essendo $I(F)$ l'informazione secondo Fisher.

La varianza asintotica rappresenta, in pratica, il prezzo da pagare, per lo stimatore scelto, quando la funzione F è la reale distribuzione sottostante gli errori. Huber trova che la varianza del "gross error model" ha un minimo e la corrispondente funzione di distribuzione di probabilità F (detta "least favorable distribution") è normale al centro ed esponenziale nelle code. Il relativo stimatore, noto come "minimax" è definito dall'espressione:

$$\psi(x) = \max(-k, \min(k, x)) \quad k > 0$$

Si tratta di uno stimatore non discendente, di tipo M , funzione del valore assegnato k (motivo per il quale occorre fissare un parametro di scala, legato alla dispersione dei dati). Si noti che lo stimatore fornisce, per $\varepsilon \rightarrow 0$ e $k \rightarrow \infty$, la media aritmetica, come pure, per $\varepsilon \rightarrow 0$ e $k \rightarrow 0$, la mediana, cioè gli stimatori di massima verosimiglianza dei casi normale ed esponenziale.

Si noti, altresì, che altre classi di stimatori maggiormente robusti, detti stimatori di tipo S , sono stimatori discendenti. Essi permettono di trattare, più correttamente e realisticamente, dati anomali, contenenti non solo piccole violazioni rispetto al modello teorico, ma anche "outliers" distanti, come i "blunders" e "leverage points".

B.2 L'approccio di Hampel

Merito fondamentale di Hampel è quello di aver introdotto uno strumento efficace per la definizione ed analisi delle proprietà di robustezza degli stimatori: la funzione di influenza. La trattazione seguente riguarda il caso monodimensionale: gli stessi concetti possono essere estesi a più dimensioni, ovviamente, con qualche ulteriore complicazione.

I tre concetti fondamentali dell'approccio infinitesimale alla robustezza sono la robustezza qualitativa, 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 tre 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 γ è un valore finito, ovvero IF è limitata, lo stimatore è robusto (tanto più robusto, quanto minore è γ).

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.

Infine il concetto di "local shift sensitivity" λ permette di comprendere, come si comporta lo stimatore per piccole variazioni di dati in una qualunque zona. Si definisce perciò:

$$\lambda = \frac{\sup_{x \neq y} |IF(y, T, F) - IF(x, T, F)|}{|y - x|}$$

intendendo, così, l'estremo superiore delle differenze (standardizzate) della IF valutata in due punti diversi. Questo corrisponde a verificare la stabilità dello stimatore, quando nel campione si tolga x ed aggiunga y , con x ed y prossimi.

L'approccio infinitesimale di Hampel è piuttosto complesso, poiché coinvolge concetti di analisi funzionale (in particolare, nel trattare gli stimatori come funzionali nello spazio delle funzioni di distribuzione di probabilità e nella definizione di una metrica, adatta ad esprimere la vicinanza di due funzioni di distribuzione di probabilità) e di teoria della stima (in particolare, nello studio delle proprietà asintotiche). Quanto segue ha, principalmente, lo scopo di fornire un quadro euristico, senza entrare in troppi dettagli.

I. Robustezza qualitativa

La definizione di robustezza qualitativa corrisponde al concetto intuitivo di robustezza, come stabilità o continuità dello stimatore, per piccoli cambiamenti nella funzione di distribuzione di probabilità sottostante.

Più precisamente, dati una variabile casuale campionaria X_n , le cui componenti siano indipendenti,

identicamente distribuite, con funzione di distribuzione di probabilità F ed uno stimatore T_n , si vuole evidenziare, come ad una piccola variazione nella funzione di distribuzione di probabilità, si avrà, in corrispondenza, una piccola variazione nella funzione di distribuzione di probabilità $L_F(T_n)$ della variabile casuale campionaria T_n . Questo fatto è, generalmente, vero per ogni n e per tutti gli stimatori ordinari, ma la continuità uniforme rispetto ad n non è garantita per gli stimatori non robusti.

La garanzia di questa stabilità di comportamento, per n crescente, può essere espressa, definendo una metrica d^* nello spazio delle misure di probabilità e richiedendo che, per $\varepsilon > 0$, esista in corrispondenza $\delta > 0$, tale che, per $n \geq n_0$, si abbia:

$$d^*(F, G) > \delta \rightarrow d^*(L_F(T_n), L_G(T_n)) < \varepsilon$$

essendo F e G due funzioni di distribuzione di probabilità vicine, come pure $L_F(T_n)$ e $L_G(T_n)$ le funzioni di distribuzione di probabilità della statistica T_n , rispettivamente, sotto F e G .

II. Funzione di influenza

Si consideri una variabile casuale campionaria X_n le cui componenti siano indipendenti, identicamente distribuite in qualche spazio campionario H . Un modello parametrico è una famiglia di funzioni di distribuzione di probabilità $F(\theta)$, essendo θ un parametro incognito, appartenente ad un qualche insieme convesso.

Allora dato un campione X_n con distribuzione empirica F_n , essendo F l'effettiva funzione di distribuzione di probabilità sottostante le osservazioni, si consideri come stimatore di θ una funzione T_n . Sia T_n consistente in F , il suo dominio un insieme convesso e si supponga T_n asintoticamente normale; si dice che per T esiste il differenziale di Gateaux, se esiste una funzione reale: $a(x)$, tale che, per ogni F , risulti:

$$\lim_{t \rightarrow 0} \frac{T((1-t)G + tH) - T(G)}{t} = \int a(x) dF(x)$$

ovvero:

$$\frac{\partial}{\partial t} [T((1-t)G + tH)]_{t=0} = \int a(x) dF(x)$$

Quest'ultima relazione esprime la derivata direzionale di T , calcolata in G , lungo la direzione di H . Si noti che è necessario esprimere l'incremento della variabile indipendente come combinazione lineare delle due funzioni di distribuzione di probabilità, perché occorre preservare il fatto che l'argomento di T sia,

comunque, una funzione di distribuzione di probabilità.

Si supponga, poi, G una funzione di distribuzione di probabilità tutta concentrata sul valore x : si dice, per definizione, che la funzione di influenza IF del funzionale T , applicata ad F , è data da:

$$IF(x, T, F) = \lim_{t \rightarrow 0} \frac{T((1-t)G + tH) - T(G)}{t}$$

per tutti gli $x \in H$ in cui tale limite esiste.

La funzione di influenza descrive l'effetto sullo stimatore T , di una contaminazione infinitesimale t nel punto x , standardizzato dalla massa t della contaminazione.

Si può, anche, dimostrare che, con qualche altra approssimazione, fra varianza asintotica e funzione d'influenza vale la relazione:

$$V(T, F) = E\left(\left(IF(x, T, F)\right)^2\right)$$

nell'ipotesi di asintotica normalità di T_n , già espressa in precedenza.

III. Punto di rottura

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 .

E' altrettanto importante, tuttavia, 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 su θ . 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) θ . 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 ε risulta, proprio, la frazione massima di "outliers" ammessa nel campione.

B.3 Il metodo della minima mediana degli scarti quadratici

Il metodo della minima mediana degli scarti quadratici avrebbe potuto chiamarsi, altrettanto bene, metodo della minima mediana degli scarti assoluti. Infatti minimizzare la mediana degli scarti assoluti significa

minimizzare il mav, cioè il valore assoluto mediano di detti scarti, operazione coincidente, completamente, con quella enunciata nel titolo del metodo proposto, essendo monotona crescente la trasformazione tra scarti assoluti e scarti quadratici. D'altra parte, i nomi sono nomi ed è inutile discuterli eccessivamente.

Il pregio di detto metodo è un punto di rottura molto alto. Infatti mentre, come noto, il punto di rottura del metodo dei minimi quadrati è nullo: basta un solo errore grossolano a distorcere una stima, il punto di rottura delle procedure robuste più classiche (ad esempio, il metodo antico della minima somma dei moduli e quello ormai classico dello stimatore di Huber) non supera il 25%. Pertanto essi sono ampiamente sufficienti per trattare insiemi di dati affetti da un numero non troppo elevato di dati anomali, ma mostrano evidenti difficoltà, quando il numero di questi cresce notevolmente, ad esempio, perché un insieme di dati è frammisto ad un altro insieme di dati, nello specifico contesto, considerati anomali.

In questi casi, di conseguenza, occorre far uso di un metodo con un altissimo punto di rottura, cosicché tutti i dati anomali possano essere evidenziati ed eliminati, conservando quella minoranza di dati, forse addirittura piccola (cioè proprio il contrario del grosso dei dati), che permette di ottenere stime non distorte. E' allora evidente come, solo per altra via, debbano essere noti e fissati i punti di taglio, in base ai quali giudicare della rispondenza, o meno, alle attese delle stime effettuate.

Il metodo della minima mediana degli scarti quadratici (ovvero della minima mediana degli scarti assoluti) cerca, in un insieme dato di m osservazioni, un sottoinsieme di equazioni, in numero identico al numero n di incognite, corrispondenti parametri del problema in esame, tali che il mav del quadrato degli scarti (o degli scarti assoluti) delle rimanenti $m - n$ sia minimo. Come evidente, la soluzione di questo problema richiede di effettuare una partizione nell'insieme delle osservazioni, utilizzandone una parte (pari al numero dei parametri) per il calcolo della soluzione e disponendo la parte rimanente come misura della bontà della soluzione trovata.

$$y = Ax + \delta \quad (P) \quad \Rightarrow \quad \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} = \begin{bmatrix} y_1^0 \\ y_2^0 \end{bmatrix} + \begin{bmatrix} 0 \\ \hat{v}_2 \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \hat{x} + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \quad \left(\begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} \right)$$

essendo: $\dim(y_1) = \dim(x) = n$ e $\dim(y_2) = m - n$

Infatti la prima partizione, dove il numero delle osservazioni è identico al numero delle incognite, ammette una ed una sola soluzione, e non dà origine ad alcun vettore di scarti-residui. Al contrario, la seconda partizione, viceversa, estranea alla soluzione trovata, ha scarti-residui e, come già detto in precedenza, serve come misura della bontà della soluzione stessa. Tanto più piccolo è il mav, calcolati su detti scarti-residui, e tanto migliore è la soluzione trovata; tanto più grande è invece lo stesso mav e tanto peggiore è allora la stessa.

Resta da osservare, come il primo gruppo di pesi P_1 sia inutile, completamente, nel calcolo della soluzione, per l'unicità della soluzione stessa, mentre il secondo gruppo di pesi P_2 , da annullare, se le osservazioni sono inserite, in modo fittizio, nel calcolo della soluzione, serva poi nella sua veste originaria, quando si tratti di ordinare i quadrati pesati degli scarti-residui (od i moduli degli stessi moltiplicati per la radice quadrata dei loro pesi).

Per quanto riguarda il calcolo della soluzione, occorre trovare, all'interno dell'insieme di m equazioni, quel sottoinsieme di n equazioni (come già detto in precedenza, pari al numero n delle incognite del sistema

dato), tali che il mav degli scarti-residui assoluti delle restanti $m - n$ equazioni sia minimo. Purtroppo, allo stato attuale delle conoscenze, non è nota altra via alternativa, a provare tutte le combinazioni possibili, secondo la nota legge combinatoria, espressa dal binomio di Newton:

$$\binom{m}{n} = \frac{m!}{n!(m-n)!} \qquad \ln \binom{m}{n} = \ln m! - \ln n! - \ln(m-n)!$$

dove i fattoriali, per i numeri grandi, possono essere facilmente calcolati con l'approssimazione di Stirling:

$$k! = \sqrt{2\pi k} \left(\frac{k}{e}\right)^k$$

Come evidente, il numero di combinazioni possibili cresce molto rapidamente. Infatti esso è 252 per 10 equazioni e 5 incognite, ma già 184.756 per 20 e 10, ben 155.117.520 per 30 e 15, ed addirittura 137.846.528.820 per 40 e 20 (si ricordi infatti che la situazione più sfavorevole, riguardo la crescita del numero delle combinazioni, si ha quando le incognite sono la metà delle equazioni).

Tuttavia, pur senza voler nascondere la gravosità del metodo, una volta risolto un primo qualsiasi sistema di n equazioni in n incognite, basta aggiornare la soluzione ottenuta, dapprima aggiungendo, ad essa, una qualsiasi delle equazioni esterne (in pratica, ripristinando il corrispondente peso p_2) e successivamente eliminando, da essa, una qualsiasi delle equazioni interne (in pratica, annullando il corrispondente peso p_1). Allora opportuni modelli partizionati ed algoritmi complementari di statistica computazionale permettono di procedere sequenzialmente, con un notevole risparmio nel tempo di calcolo.

Inoltre quando il sistema dato presenta equazioni che non coinvolgono tutte le incognite, lo scambio fra un'equazione ed un'altra è strettamente legato, alla capacità della seconda di svolgere un ruolo simile, se non identico, alla prima, pena scrivere un sistema singolare, perché qualche incognita cessa di essere indirettamente osservata dalle equazioni. Allora il numero delle combinazioni possibili tende a decrescere, pari alla somma di tutte le combinazioni possibili su sottoinsiemi di osservaioni compatibili fra loro, anche se la loro individuazione può non essere affatto semplice.

Infine resta da osservare, come trattandosi di un metodo sofisticato per l'individuazione di dati anomali, sia sempre possibile applicare lo stesso solo su ben selezionati sottoinsiemi omogenei di dati, di dimensione ragionevole, ancora una volta, contenendo il numero di combinazioni possibili, al massimo.

APPENDICE C – MODELLI PARTIZIONATI ¹⁵

C.1 Il filtro di Kalman

I modelli partizionati e, fra questi, principalmente, il filtro di Kalman ed i blocchi di Gauss (di cui sarà detto nel prosieguo), servono a risolvere problemi ai minimi quadrati, senza ricalcolare ex-novo l'intera soluzione e la sua matrice di varianza-covarianza, quando nel problema in esame intervenga una qualche variazione nel

¹⁵ Questa appendice riprende ed aggiorna quanto esposto in: Bellone T., Mussio L. (1994): Algoritmi esatti sequenziali. In: F. Crosilla, L. Mussio (Ed's) *Metodi e procedure avanzate di modellizzazione e trattamento dei dati GPS*. CISM, Udine, p. 67-109.

numero delle equazioni e/o nel numero delle incognite, a seguito dell'immissione od eliminazione di osservazioni e/o parametri nel sistema dato.

Dato un problema ai minimi quadrati, nella usuale forma di Gauss–Markov (con equazioni d'osservazione e/o pseudo–osservazione sovrappesate e non), l'algoritmo del filtro di Kalman risolve un sistema, prendendo in considerazione una soluzione di un sottoinsieme di equazioni. Lo stesso metodo è usato per aggiornare la matrice di varianza–covarianza.

$$E \left(\begin{array}{c} y_1 \\ y_2 \end{array} \right) = \begin{array}{c} \overline{y_1} \\ \overline{y_2} \end{array} = \begin{array}{c} A \\ B \end{array} \begin{array}{c} | \\ | \\ x \end{array}$$

$$D \left(\begin{array}{c} y_1 \\ y_2 \end{array} \right) = \begin{array}{cc} C_{y_0_1 y_0_1} & 0 \\ 0 & C_{y_0_2 y_0_2} \end{array} = \sigma_0^2 \begin{array}{cc} Q_{y_0_1 y_0_1} & 0 \\ 0 & Q_{y_0_2 y_0_2} \end{array} = \sigma_0^2 \begin{array}{cc} P^{-1} & 0 \\ 0 & Q^{-1} \end{array}$$

Se è disponibile la soluzione ai minimi quadrati di una partizione del problema stesso:

$$\hat{x}_0 = (A^T P A)^{-1} A^T P y_{0_1}$$

corredata dalla sua matrice di varianza–covarianza, ottenuta al solito applicando la legge di propagazione della covarianza:

$$C_{\hat{x}_0 \hat{x}_0} = \sigma_0^2 Q_{\hat{x}_0 \hat{x}_0} = \sigma_0^2 (A^T P A)^{-1}$$

la soluzione ai minimi quadrati dell'intero problema:

$$\hat{x} = (A^T P A + B^T Q B)^{-1} (A^T P y_{0_1} + B^T Q y_{0_2})$$

corredata dalla sua matrice di varianza–covarianza:

$$C_{\hat{x} \hat{x}} = \sigma_0^2 Q_{\hat{x} \hat{x}} = \sigma_0^2 (A^T P A + B^T Q B)^{-1}$$

può essere ottenuta in funzione della soluzione precedente e di un'innovazione.

E' facile rilevare, come lo stesso algoritmo possa essere ripercorso "a gambero", partendo da una soluzione più generale ed arrivando ad una sua sottosoluzione, sempre corredata della sua matrice di varianza–covarianza.

Dato infatti un intero problema ai minimi quadrati, nella usuale forma di Gauss – Markov, del quale si voglia conoscere una sua opportuna riduzione:

$$E(y_1) = E \left(\begin{pmatrix} \bullet \\ y_2 \end{pmatrix} \right) = \overline{y_1} = \begin{pmatrix} \bullet \\ \overline{y_2} \end{pmatrix} = A x_0 = \begin{pmatrix} \bullet \\ B \end{pmatrix} x_0$$

$$D(y_1) = D \left(\begin{pmatrix} \bullet \\ y_2 \end{pmatrix} \right) = C_{y_0_1 y_0_1} = \begin{vmatrix} \bullet & 0 \\ 0 & C_{y_0_2 y_0_2} \end{vmatrix} = \sigma_0^2 Q_{y_0_1 y_0_1} = \sigma_0^2 \begin{vmatrix} \bullet & 0 \\ 0 & Q_{y_0_2 y_0_2} \end{vmatrix} = \sigma_0^2 P^{-1} = \sigma_0^2 \begin{vmatrix} \bullet & 0 \\ 0 & Q^{-1} \end{vmatrix}$$

Se è disponibile la soluzione ai minimi quadrati dell'intero problema:

$$\hat{x}_0 = (A^T P A)^{-1} A^T P y_{0_1}$$

corredata della sua matrice di varianza – covarianza:

$$C_{\hat{x}_0 \hat{x}_0} = \sigma_0^2 Q_{\hat{x}_0 \hat{x}_0} = \sigma_0^2 (A^T P A)^{-1}$$

la soluzione ai minimi quadrati del problema ridotto:

$$\hat{x} = (A^T P A - B^T Q B)^{-1} (A^T P y_{0_1} - B^T Q y_{0_2})$$

corredata dalla sua matrice di varianza–covarianza:

$$C_{\hat{x} \hat{x}} = \sigma_0^2 Q_{\hat{x} \hat{x}} = \sigma_0^2 (A^T P A - B^T Q B)^{-1}$$

può essere ottenuta in funzione della soluzione precedente e di un'innovazione "a gambero".

Infatti data la soluzione ai minimi quadrati dell'intero problema, effettuate le debite sostituzioni ed applicando il noto lemma per l'inversione di una matrice, si ottiene:

$$\begin{aligned} \hat{x} &= (A^T P A \pm B^T Q B)^{-1} (A^T P y_{0_1} \pm B^T Q y_{0_2}) = \\ &= (Q_{\hat{x}_0 \hat{x}_0}^{-1} \pm B^T Q_{y_0_2 y_0_2}^{-1} B)^{-1} (A^T P y_{0_1} \pm B^T Q y_{0_2}) = \\ &= (Q_{\hat{x}_0 \hat{x}_0} \mp Q_{\hat{x}_0 \hat{x}_0} B^T (Q_{y_0_2 y_0_2} \pm B Q_{\hat{x}_0 \hat{x}_0} B^T)^{-1} B Q_{\hat{x}_0 \hat{x}_0}) (A^T P y_{0_1} \pm B^T Q y_{0_2}) \end{aligned}$$

Dopodiché distribuita la somma algebrica rispetto al prodotto, effettuate le debite sostituzioni ed applicato di nuovo il lemma per l'inversione di una matrice, limitatamente al terzo addendo (avendo anche raccolto a fattore comune, davanti alla parentesi interna, i segni: \pm , e notando che si ha sempre: $\mp \cdot \pm = -$), risulta:

$$\begin{aligned}
\hat{x} &= Q_{\hat{x}_0, \hat{x}_0} A^T P y_{0_1} \mp Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} B Q_{\hat{x}_0, \hat{x}_0} A^T P y_{0_1} + \\
&\pm \left(Q_{\hat{x}_0, \hat{x}_0} - Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} B Q_{\hat{x}_0, \hat{x}_0} \right) B^T Q y_{0_2} = \\
&= \hat{x}_0 \mp Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} B \hat{x}_0 \pm \left(Q_{\hat{x}_0, \hat{x}_0} - Q_{\hat{x}_0, \hat{x}_0} B^T \left(B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} + \right. \\
&\left. \mp \left(B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} \left(Q_{y_{0_2}, y_{0_2}}^{-1} \pm \left(B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} \right)^{-1} \left(B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} B Q_{\hat{x}_0, \hat{x}_0} \right) B^T Q_{y_{0_2}, y_{0_2}}^{-1} y_{0_2}
\end{aligned}$$

Allora sempre limitatamente al terzo addendo, effettuato il raccoglimento a fattore comune ($Q_{\hat{x}_0, \hat{x}_0} B^T$ a sinistra e $B Q_{\hat{x}_0, \hat{x}_0} B^T$ a destra) ed applicate, nell'ordine, le identità:

$$A^{-1} B^{-1} C^{-1} = (CBA)^{-1} \quad \text{e} \quad A(A^{-1} \pm B^{-1})B = B \pm A = \pm(A \pm B) \quad \text{si ha:}$$

$$\begin{aligned}
\hat{x} &= \hat{x}_0 \mp Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} \hat{y}_2 \pm Q_{\hat{x}_0, \hat{x}_0} B^T \left(I + \right. \\
&\left. - \left(I \mp \left(B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} \left(Q_{y_{0_2}, y_{0_2}}^{-1} \pm \left(B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} \right)^{-1} \right) \left(B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} B Q_{\hat{x}_0, \hat{x}_0} B^T \right) Q_{y_{0_2}, y_{0_2}}^{-1} y_{0_2} = \\
&= \hat{x}_0 \mp Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} \hat{y}_2 + \\
&\pm Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \left(Q_{y_{0_2}, y_{0_2}}^{-1} \pm \left(B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} \right) B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} y_{0_2} = \\
&= \hat{x}_0 \mp Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} \hat{y}_2 \pm Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} y_{0_2}
\end{aligned}$$

(avendo anche tenuto conto, quattro volte, della regola dei segni, ovvero: $\pm \cdot - \cdot \mp \cdot \pm = \pm$).

Infine effettuato di nuovo il raccoglimento a fattore comune ed introdotta la matrice di amplificazione H , la soluzione dell'intero problema può essere espressa come la somma (o la differenza) della soluzione precedente e dell'innovazione ("a gambero") \hat{w} :

$$\hat{x} = \hat{x}_0 \pm Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm Q_{\hat{y}_2, \hat{y}_2} \right)^{-1} (y_{0_2} - \hat{y}_2) = \hat{x}_0 \pm H \hat{v} = \hat{x}_0 \pm \hat{w}$$

dove l'innovazione è data dal prodotto della matrice di amplificazione per la discrepanza \hat{w} fra le nuove osservazioni y_{0_2} ed i valori predetti dalla soluzione precedente \hat{y}_2 .

Analogamente data la matrice di varianza-covarianza, effettuate le debite sostituzioni ed applicando il lemma per l'inversione di una matrice (di Faddejeva), si ottiene:

$$\begin{aligned}
C_{\hat{x}\hat{x}} &= \sigma_0^2 \left(A^T P A \pm B^T Q B \right)^{-1} = \sigma_0^2 \left(Q_{\hat{x}_0, \hat{x}_0}^{-1} \pm B^T Q_{y_{0_2}, y_{0_2}}^{-1} B \right)^{-1} = \\
&= \sigma_0^2 Q_{\hat{x}_0, \hat{x}_0} \mp \sigma_0^2 Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm B Q_{\hat{x}_0, \hat{x}_0} B^T \right)^{-1} B Q_{\hat{x}_0, \hat{x}_0} = \\
&= \sigma_0^2 Q_{\hat{x}_0, \hat{x}_0} \mp \sigma_0^2 Q_{\hat{x}_0, \hat{x}_0} B^T \left(Q_{y_{0_2}, y_{0_2}} \pm Q_{\hat{y}_2, \hat{y}_2} \right)^{-1} B Q_{\hat{x}_0, \hat{x}_0} = \sigma_0^2 Q_{\hat{x}_0, \hat{x}_0} \mp \sigma_0^2 Q_{\hat{x}_0, \hat{x}_0} B^T Q_{\hat{v}\hat{v}}^{-1} B Q_{\hat{x}_0, \hat{x}_0}
\end{aligned}$$

Dopodiché trasformata l'espressione appena trovata nella forma classica della legge di propagazione di covarianza ed introdotta la matrice di amplificazione H (o di guadagno, dall'inglese: *gain*), la matrice di varianza-covarianza può essere espressa come differenza (somma) fra la matrice di varianza-covarianza della soluzione precedente e la matrice di varianza-covarianza dell'innovazione ("a gambero"):

$$C_{\hat{x}\hat{x}} = \sigma_0^2 Q_{\hat{x}_0\hat{x}_0} \mp \sigma_0^2 Q_{\hat{x}_0\hat{x}_0} B^T Q_{\hat{v}\hat{v}}^{-1} Q_{\hat{v}\hat{v}} Q_{\hat{v}\hat{v}}^{-1} B Q_{\hat{x}_0\hat{x}_0} = \sigma_0^2 Q_{\hat{x}_0\hat{x}_0} \mp \sigma_0^2 H Q_{\hat{v}\hat{v}} H^T = \sigma_0^2 Q_{\hat{x}_0\hat{x}_0} \mp \sigma_0^2 Q_{\hat{w}\hat{w}}$$

Si osservi come la matrice di varianza-covarianza della soluzione somma (differenza), per l'introduzione di nuove osservazioni completamente indipendenti (l'eliminazione di vecchie osservazioni ovviamente non indipendenti) è data dalla differenza fra la matrice di varianza-covarianza della soluzione precedente e la matrice di varianza-covarianza dell'innovazione ("a gambero").

C.2 I Blocchi di Gauss

L'algoritmo dei blocchi di Gauss risolve un sistema lineare, prendendo in considerazione una soluzione di un sottoinsieme di equazioni, note in precedenza. Lo stesso metodo è usato per aggiornare la matrice inversa del suddetto sistema lineare. Pertanto dato un sistema lineare:

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

una soluzione di un suo sottoinsieme di equazioni:

$$x_{0_1} = A^{-1} y_1$$

del quale sia stata calcolata anche la sua matrice inversa, calcolati gli elementi della sua matrice inversa del suddetto sistema lineare (a partire dalla matrice inversa del sottoinsieme di equazioni):

$$\left| \begin{array}{cc|cc} A & B & \alpha & \beta \\ B^T & C & \beta^T & \gamma \end{array} \right| = \left| \begin{array}{cc|cc} I & 0 \\ 0 & I \end{array} \right|$$

$$\begin{cases} \alpha = A^{-1} + A^{-1} B (C - B^T A^{-1} B)^{-1} B^T A^{-1} = \\ \quad = A^{-1} + A^{-1} B \gamma B^T A^{-1} \\ \beta = -A^{-1} B (C - B^T A^{-1} B)^{-1} = -A^{-1} B \gamma \\ \gamma = (C - B^T A^{-1} B)^{-1} \end{cases} \quad \Rightarrow \quad A^{-1} B = -\beta \gamma^{-1}$$

si ottiene:

$$\left| \begin{array}{c} x_1 \\ x_2 \end{array} \right| = \left| \begin{array}{cc|c} \alpha & \beta & y_1 \\ \beta^T & \gamma & y_2 \end{array} \right| = \left| \begin{array}{c} A^{-1} y_1 + A^{-1} B (C - B^T A^{-1} B)^{-1} (B^T A^{-1} y_1 - y_2) \\ -(C - B^T A^{-1} B)^{-1} (B^T A^{-1} y_1 - y_2) \end{array} \right| = \left| \begin{array}{c} x_{0_1} - A^{-1} B y_2 \\ -(C - B^T A^{-1} B)^{-1} (B^T x_{0_1} - y_2) \end{array} \right|$$

ovvero la soluzione del suddetto sistema lineare, data una sua sottosoluzione nota in precedenza.

Infine data la soluzione di un sistema lineare:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

del quale sia stata calcolata anche la sua matrice inversa, le stesse relazioni permettono anche il calcolo di una sua sottosoluzione:

$$x_{0_1} = x_1 + A^{-1} B y_2 = x_1 - \beta \gamma^{-1} y_2$$

sempre corredata dalla matrice inversa di quest'ultima:

$$A^{-1} = \alpha - A^{-1} B \gamma B^T A^{-1} = \alpha - \beta \gamma^{-1} \beta^T$$

prendendo in considerazione rispettivamente la soluzione e la matrice inversa dell'intero sistema 16.

¹⁶ Una nota conclusiva, sui modelli partizionati, mette in evidenza come il calcolo della soluzione di grandi sistemi sia un'operazione computazionalmente lunga e complessa. Pertanto ogni qualvolta occorre calcolarne una modifica è opportuno procedere in modo sequenziale, ovvero modificare la soluzione già ottenuta in precedenza, invece di ripetere ex-novo il calcolo della soluzione.

Alcune varianti degli algoritmi di Householder, Givens e Cholesky (cui si rimanda) consentono il calcolo sequenziale della soluzione, ovvero di aggiungere o togliere da essa una o più equazioni alle misure oppure una o più incognite. Queste operazioni sono utili, ad es., rispettivamente, per l'eliminazione delle equazioni alle misure corrispondenti ad osservazioni affette da errori grossolani e per l'aggiornamento nel modello funzionale dei parametri (aggiuntivi) incogniti correttivi di alcuni errori sistematici o pseudo-sistematici, la cui presenza si sospetta nell'insieme delle misure, in base alla pratica sperimentale. Inoltre l'adeguata applicazione di un noto teorema dell'algebra delle matrici consente il calcolo sequenziale della matrice di varianza-covarianza delle incognite.

Si osservi poi che una variazione del peso di un'equazione alle misure, come nella stima contemporanea delle incognite e dei pesi delle quantità osservate (effettuata con stime di tipo Helmert, ovvero con un algoritmo iterativo detto di riproduzione dei pesi) e nei problemi di ottimizzazione dei pesi delle quantità osservate, altrimenti detti di secondo ordine (SOD), corrisponde all'aggiungere o al togliere un'equazione alle misure, pesata tanto quanto l'aumento o la diminuzione di peso che si deve apportare. Per quanto riguarda invece i problemi di ottimizzazione dello schema di misura, altrimenti detti di primo ordine (FOD), se eseguiti iterativamente per mezzo dell'ottimizzazione di secondo ordine con arresto in base al controllo sull'affidabilità delle quantità osservate, questi comportano di aggiungere o di togliere un'equazione alle misure. Sempre una variazione del peso di un'equazione alle misure si presenta con l'uso di procedure robuste (quali la minima somma dei moduli, oppure i criteri di Huber o di Hampel), dove il calcolo della soluzione è ottenuto tramite la ripetizione ripesata di una compensazione minimi quadrati.

Ancora le stesse operazioni sono utili nei problemi in cui occorre aggiungere o togliere un punto e tutte le misure che lo determinano, oppure aggiungere o togliere un insieme di punti in un diverso sistema di riferimento e anche le misure relative a tutti i punti ivi compresi. In caso di introduzione, dapprima si aggiungono al sistema, equazione per equazione, tutti i coefficienti delle nuove equazioni relativi alle vecchie incognite e successivamente si aggiungono al sistema così modificato, un'incognita alla volta, tutti i coefficienti sempre delle nuove equazioni relative alle nuove incognite. Invece in caso di esclusione, dapprima si tolgono dal sistema, una incognita alla volta, tutti i coefficienti delle equazioni da eliminare relativi alle incognite da eliminare e successivamente si tolgono dal sistema così modificato, equazione per equazione, tutti i coefficienti delle equazioni da eliminare relativi alle incognite da conservare.

Infine per l'interpolazione ad elementi finiti di un campione di dati con un insieme di funzioni splines specializzate rispettivamente per trattare i dati di una serie temporale (o altro campione monodimensionale) oppure quelli di un lattice a supporto territoriale o spaziale, nel caso in cui passi e finestre siano variabili da zona a zona, si opera pure sequenzialmente. In questo caso, tramite l'analisi di varianza (confronto fra la varianza spiegata e la varianza residua), si procede al campionamento progressivo, cioè all'acquisizione di nuove osservazioni, nelle zone dove l'approssimazione non è accettabile, mentre si procede al campionamento selettivo, cioè al diradamento delle osservazioni, nelle zone dove l'approssimazione è più che accettabile; questo approccio è particolarmente indicato con l'impiego di tecniche avanzate multirisoluzione e multilivello.

Gli algoritmi sono predisposti per aggiungere o per togliere, ogni volta, una sola equazione alle misure o una sola incognita; pertanto nel caso in cui si devono aggiungere o togliere più equazioni alle misure o più incognite, occorre ripetere l'applicazione di ciascun algoritmo richiesto tante volte quante sono le equazioni alle misure o le incognite da aggiungere o togliere.

Il vantaggio dell'uso di questi algoritmi consiste nella possibilità di modificare con essi i risultati già ottenuti con un numero di operazioni dell'ordine del quadrato del numero delle incognite, anziché del cubo delle stesse, come si avrebbe utilizzando gli algoritmi di base. Si noti perciò che anche l'aggiungere o il togliere più (poche) equazioni alle misure o più (poche) incognite è ancora un'operazione abbastanza sollecitata.

La ragione dell'esposizione a parte degli algoritmi sequenziali, rispetto agli algoritmi di base da cui discendono, sta nel fatto che, come appena osservato, ben differenti sono le finalità del loro utilizzo. Infatti mentre dal punto di vista dei metodi di calcolo numerico il capitolo algoritmi diretti deve considerarsi unico, dal punto di vista della loro applicazione, nello specifico, alle scienze geodetiche e geomatiche, una cosa è la compensazione minimi quadrati di un grande numero di osservazioni ed un'altra è l'eliminazione degli errori grossolani o la correzione degli errori sistematici o pseudo-sistematici e così via. Pertanto poiché come è evidente, si tratta di problematiche distinte, esporre insieme gli algoritmi di base e gli algoritmi sequenziali avrebbe contribuito a confondere anziché a chiarire.

APPENDICE D – DUE SEMPLICI ESEMPI NUMERICI

Due semplici esempi numerici svolgono passo, passo un problema a minimi quadrati nel caso principe di osservazioni di condizione con parametri. Il primo esempio tratta l'interpolazione lineare di un qualsiasi fenomeno distribuito nello spazio 3D, dove il modello funzionale è ovviamente costituito da un iperpiano ed il modello stocastico è semplicemente la matrice identità, in quanto si è scelto di considerare tutte le osservazioni indipendenti tra loro ed aventi la stessa precisione.

INTERPOLAZIONE LINEARE IN UN DOMINIO SPAZIALE

$$S = A + BX + CY + DZ + H \quad \text{essendo: } H = -0.60, \text{ una costante assegnata }^{17}.$$

DATI IN INGRESSO

VAR. IND. X	VAR. IND. Y	VAR. IND. Z	S OSSERVATO
0.71	0.62	0.17	-1.50
-0.71	0.62	0.17	0.11
0.71	-0.62	0.17	0.57
0.71	0.62	-0.17	-0.35
-0.71	-0.62	0.17	0.26
-0.71	0.62	-0.17	-0.06
0.71	-0.62	-0.17	-0.55
-0.71	-0.62	-0.17	-0.54

NUMERO DI OSSERVAZIONI ED EQUAZIONI = 8

NUMERO DI PARAMETRI ED INCOGNITE = 4

ELABORAZIONI

MATRICE DISEGNO A ¹⁸

1.00	0.71	0.62	0.17
1.00	-0.71	0.62	0.17
1.00	0.71	-0.62	0.17
1.00	0.71	0.62	-0.17
1.00	-0.71	-0.62	0.17
1.00	-0.71	0.62	-0.17
1.00	0.71	-0.62	-0.17
1.00	-0.71	-0.62	-0.17

TERMINE NOTO B

0.90
-0.71
-1.17
-0.25
-0.86
-0.54
-0.05
-0.06

MATRICE NORMALE C

8.00	0.00	0.00	0.00
	4.03	0.00	0.00
matrice		3.08	0.00
simmetrica			0.23

T. NOTO NORMALE D

-2.74
1.14
0.95
-0.16

¹⁷ L'assegnazione aggiuntiva di una costante numerica può sembrare una complicazione inutile e quasi una stravaganza; tuttavia essa obbliga a scrivere le espressioni complete delle stime, senza cancellare termini nulli, cosa molto utile quando si è costretti a prendere in considerazione modelli funzionali non – lineari, dove una costante numerica è data dallo stesso modello funzionale, calcolato in base ai valori approssimati dei parametri incogniti.

¹⁸ Si noti come non è stata scritta, qui né altrove, alcuna matrice trasposta, in quanto del tutto inutile nel prosieguo dei calcoli. Infatti l'uso per righe o per colonne di una qualsiasi matrice trasposta, è equivalente all'uso per colonne o per righe della matrice che l'ha generata e, di conseguenza, scrivere una matrice trasposta non dà alcuna informazione aggiuntiva e può al più generare qualche indesiderato errore di copiatura.

Un commento sulle elaborazioni effettuate rileva la banalità della scrittura della matrice disegno del sistema di equazioni d'osservazione che, dopo una colonna di numeri uno, riferita alla prima incognita costante, ricopia semplicemente la tabella dei dati in ingresso, per quanto riguarda le tre incognite lineari, riferite alle tre variabili indipendenti del dominio spaziale. Anche vettore il termine noto è semplicemente ottenuto facendo la differenza tra la costante assegnata ed il vettore delle osservazioni.

Dopodiché il proseguimento delle elaborazioni prevede il calcolo della matrice normale e quello del vettore termine noto normale che richiedono di effettuare:

- ❑ somme quadratiche pesate degli elementi di ciascuna colonna della matrice disegno, per ottenere tutti gli elementi della diagonale principale della matrice normale;
- ❑ somme di prodotti pesati degli elementi di tutte le coppie distinte delle colonne della matrice disegno, per ottenere tutti gli elementi esterni alla diagonale principale della matrice normale;
- ❑ somme di prodotti pesati degli elementi di ciascuna colonna della matrice disegno sempre per gli elementi del vettore del termine noto, per ottenere tutti gli elementi del vettore termine noto normale,

essendo possibile osservare che somme pesate diventano semplici somme nel caso la matrice dei pesi sia la matrice identità, come pure che basta calcolare i soli elementi al di sopra della diagonale principale della matrice normale, perché la stessa matrice è simmetrica ¹⁹.

Dopo questa operazione, è agevole ottenere i risultati del problema proposto, ovvero nell'ordine: i vettori della soluzione a minimi quadrati, delle osservazioni compensate e degli scarti – residui. Infatti la soluzione a minimi quadrati è data dal prodotto tra la matrice inversa della matrice normale ed il vettore termine noto normale, mentre le osservazioni compensate sono ottenute moltiplicando la matrice disegno per il vettore della soluzione e gli scarti – residui calcolando le differenze tra le osservazioni compensate e le osservazioni (in ingresso). La somma quadratica pesata (se necessario) degli elementi del terzo vettore fornisce uno scalare la cui radice quadrata è comunemente chiamata sigma zero.

RISULTATI

SOLUZIONE X SQM SOLUZIONE

0.34	0.26
-0.28	0.36
-0.31	0.41
0.69	1.51

SCARTI-RESIDUI V SQM SCARTI-RESIDUI S STIMATO SQM S STIMATO

0.97	0.51	-0.53	0.51
-0.24	0.51	-0.13	0.51
-0.72	0.51	-0.15	0.51
-0.42	0.51	-0.77	0.51
-0.01	0.51	0.25	0.51
-0.31	0.51	-0.37	0.51
0.17	0.51	-0.38	0.51
0.56	0.51	0.02	0.51

¹⁹ Il caso più generale di un modello stocastico comprendente correlazioni tra le osservazioni richiede invece di invertire la suddetta matrice, dovendo successivamente tenerne conto nel calcolo delle somme pesate che, di conseguenza, si fanno più complesse.

SIGMA ZERO 0.73 N. CONDIZIONE 0.03

RIDONDANZE LOCALI 0.50 0.50 0.50 0.50 0.50 0.50 0.50 0.50

Infine le radici quadrate degli elementi della diagonale principale della matrice inversa della matrice normale, moltiplicate per sigma zero, forniscono gli sqm della soluzione, mentre per ricavare gli sqm degli scarti – residui e delle osservazioni compensate, occorre procedere come di seguito riportato. Infatti le varianze delle osservazioni compensate, a meno del quadrato di sigma zero, sono calcolate mediante un prodotto scalare tra una data riga della matrice disegno e la sua trasposta, avendo per nucleo centrale la sopraccitata matrice inversa ²⁰. Da ultimo, le varianze degli scarti – residui, a meno del quadrato di sigma zero, sono calcolate dalle differenze tra l'inverso dei pesi e i corrispondenti cofattori relativi alle osservazioni compensate (cioè le loro varianze, ancora a meno del quadrato di sigma zero).

Un commento sulla documentazione di corredo che permette di eseguire i dovuti controlli numerici, sempre necessari accanto a quelli statistici, forniti da sigma zero e da tutti gli sqm calcolati, mette in evidenza come questi diano informazioni buon condizionamento dei parametri e sull'affidabilità delle osservazioni. Infatti l'inverso del prodotto delle norme degli estremi superiori delle matrici normali ed inversa fornisce il numero di condizione e dà informazioni sul buon condizionamento, mentre i prodotti dei cofattori degli scarti – residui (ovvero le loro varianze, a meno del quadrato di sigma zero) per i pesi delle corrispondenti osservazioni permettono di calcolare le ridondanze locali delle osservazioni e danno informazioni sulla loro affidabilità.

Il secondo esempio tratta l'interpolazione polinomiale di un qualsiasi processo temporale, dove il modello funzionale è ovviamente costituito da una parabola cubica ed il modello stocastico è sempre la matrice identità. Si osservi, a riguardo, come il secondo esempio, benché apparentemente meno complesso del primo, perché distribuito in uno dominio più semplice, sia invece più complesso, per il tipo di interpolazione richiesta ed alcune particolarità computazionali che ne discendono.

INTERPOLAZIONE POLINOMIALE CUBICA
 $S = A + BT^2 + CT + DT^3 + H$ essendo: $H = 1.46$.

<u>DATI IN INGRESSO</u>	
VAR. IND. T	S OSSERVATO
0.17	-1.50
-0.17	0.11
0.34	0.57
-0.34	-0.35
0.51	0.26
-0.51	-0.06
0.68	-0.55
-0.68	-0.54

NUMERO DI OSSERVAZIONI ED EQUAZIONI = 8

20 Per completezza si ricorda che il calcolo della matrice inversa è l'inverso dei singoli elementi, se la matrice è diagonale; altrimenti è l'inversa dei singoli blocchi diagonali, se la matrice è diagonale a blocchi (di cui, nei casi 2×2 e 3×3 , esistono semplici e ben note formule chiuse). Nel caso più generale, noti algoritmi di algebra lineare permettono il calcolo della soluzione (che prescinde dal calcolo della matrice inversa) e della matrice inversa (ove richiesto) e si rimanda ad essi, notando che, per matrici non troppo grandi, gli stessi algoritmi sono solitamente implementati nei programmi di calcolo scientifico e, oggigiorno addirittura, nei fogli elettronici.

NUMERO DI PARAMETRI ED INCOGNITE = 4

ELABORAZIONI

MATRICE DISEGNO A TERMINE NOTO B

1.00	0.03	0.17	0.00	2.96
1.00	0.03	-0.17	-0.00	1.35
1.00	0.12	0.34	0.04	0.89
1.00	0.12	-0.34	-0.04	1.81
1.00	0.26	0.51	0.13	1.20
1.00	0.26	-0.51	-0.13	1.52
1.00	0.46	0.68	0.31	2.01
1.00	.46	-0.68	-0.31	2.00

MATRICE NORMALE C ²¹ T. NOTO NORMALE D

8.00	1.73	0.00	0.00	13.74
	0.59	0.00	0.00	3.00
matrice		1.73	0.59	-0.20
simmetrica			0.24	-0.07

RISULTATI

SOLUZIONE X SQM SOLUZIONE

-1.70	0.49
-0.09	1.81
0.10	1.67
0.03	4.53

SCARTI-RESIDUI V SQM SCARTI-RESIDUI S STIMATO SQM S STIMATO

1.28	0.66	-0.22	0.52
-0.37	0.66	-0.26	0.52
-0.78	0.65	-0.21	0.54
0.07	0.65	-0.28	0.54
-0.47	0.69	-0.21	0.48
-0.26	0.69	-0.32	0.48
0.35	0.30	-0.20	0.78
0.18	0.30	-0.36	0.78

SIGMA ZERO 0.84 N. CONDIZIONE 0.00

RIDONDANZE LOCALI 0.61 0.61 0.59 0.59 0.67 0.67 0.13 0.13

21 Due osservazioni sono di rigore. La prima mette in evidenza che la stampa delle elaborazioni e dei risultati presenta costantemente due cifre decimali, così come sono forniti i dati in ingresso. Tuttavia è facile constatare come, per lo svolgimento corretto dei calcoli, siano necessarie almeno quattro (o cinque) cifre decimali, nella scrittura della matrice disegno, ed addirittura otto (o dieci) cifre significative, nel calcolo della matrice normale e della sua inversa. Infatti il quadrato di un numero piccolo e maggiormente il cubo dello stesso numero tendono a zero abbastanza rapidamente, senza tuttavia essere nulli, ed una rappresentazione decimale finita, con poche cifre decimali, fa apparire nulli numeri piccoli, ma non affatto nulli, di conseguenza, producendo poi risultati errati. La seconda osservazione giustifica l'ordine delle incognite; infatti raggruppando insieme gli esponenti pari (considerando pari anche lo zero) e separatamente quelli dispari, si ottiene una matrice diagonale a blocchi che altrimenti si presenta sparsa. Poiché l'algoritmica di una matrice diagonale a blocchi è più semplice di quella di una matrice sparsa, questo è certamente un motivo sufficiente per scegliere questo ordine delle incognite. Proprio da tutto ciò, risulta evidente la maggiore complessità del secondo esempio rispetto al primo.

APPENDICE E – ALCUNI ESEMPI DI RETI GEODETICHE E BLOCCHI FOTOGRAMMETRICI

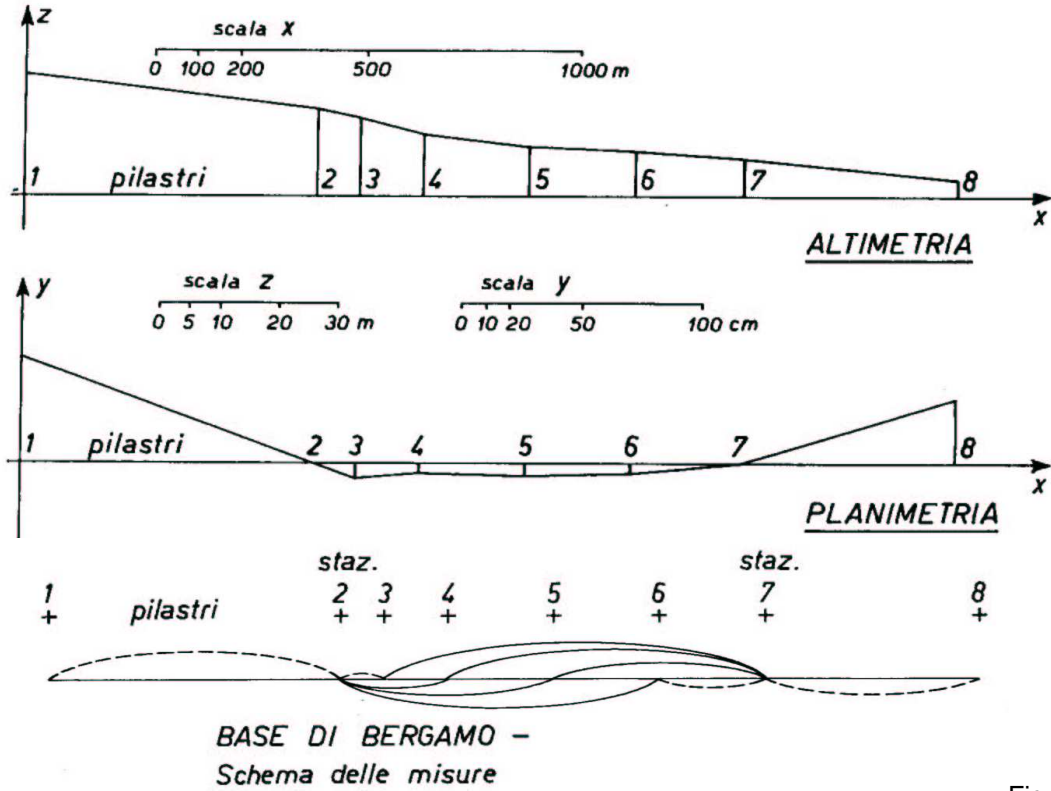
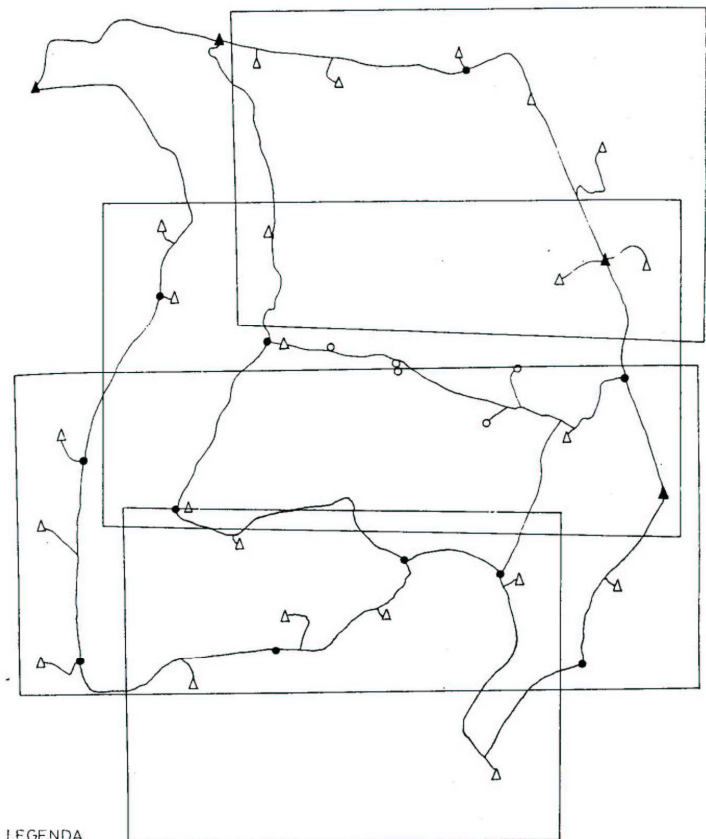


Fig. E.1

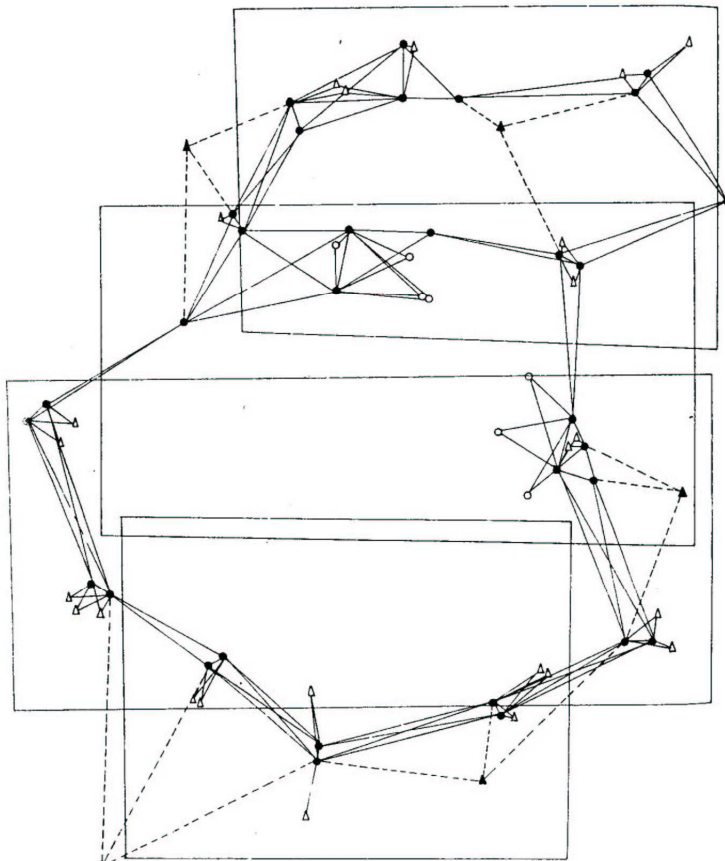


LEGENDA

- ▲ Caposaldo fondamentale
- Caposaldo
- △ Punto di appoggio
- Punto di controllo



Fig. E.2

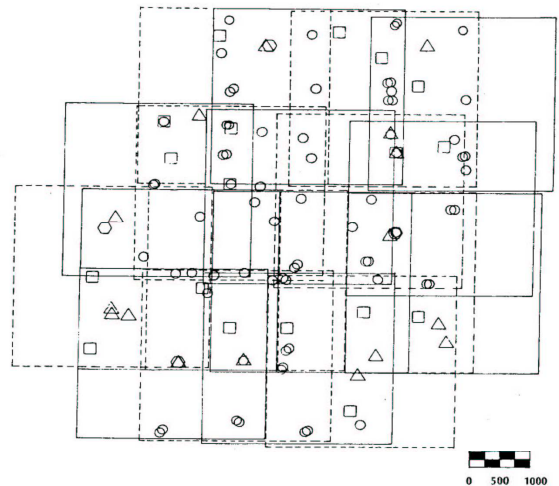


LEGENDA

- ▲ Vertice di inquadramento
- Vertice di poligonale
- △ Punto di appoggio
- Punto di controllo



Fig. E.3



LEGENDA

- Punti di legame
- △ Punti di appoggio planimetrici
- Punti di appoggio altimetrici
- Punti di appoggio piano-altimetrici

Fig. E.4 22

22 La figura E.1 mostra il profilo altimetrico, l'andamento planimetrico e lo schema di misure della base distanziometrica di precisione presso Villa d'Almè (Bergamo); le figure E.2, E.3 ed E.4 rappresentano le reti d'appoggio altimetrica e planimetrica ed il blocco fotogrammetrico di immagini, per la formazione di cartografia tecnica nella parte del Parco del Ticino tra Sesto Calende e Golasecca. I sistemi di equazioni d'osservazione delle misure effettuate per la determinazione delle incognite, costituite dalle coordinate di caposaldi, vertici e/o punti fotografici, nonché altri parametri di servizio, sono state compensate a minimi quadrati e sottoposte a test di validazione.

APPENDICE F – ANALISI DELLA CONNESSIONE, DI VARIANZA E FATTORIALE

E.1 Connessione multipla

Un piano mirato di esperimenti simulati sul comportamento di un materiale è stato programmato, in base ad una tabella di connessione multipla. Infatti detta tabella, costruita mediante il calcolo delle contingenze multiple, fornisce i dati intermedi per la stima degli indici multipli di Bonferroni, i cui valori variano tra zero ed uno, ed indicano connessione maggiore per i valori più alti (di seguito, si riportano le espressioni usate 23).

Contingenze:
$$c_{ijk} = f_{ijk} - p_i q_j r_k \quad \text{dove: } -1 \leq c_{ijk} \leq 1$$

Semi – contingenza media:
$$C_0 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^l |c_{ijk}|$$

Indici di Bonferroni unilaterali

$$\left\{ \begin{array}{l} \beta_X = \frac{C_0}{1 - \sum_{j=1}^m \sum_{k=1}^l R_{jk}^3} \quad ; \text{ indipendenza } 0 \leq \beta_X \leq 1 \text{ perfetta dipendenza } x = h(yz) \\ \beta_Y = \frac{C_0}{1 - \sum_{i=1}^n \sum_{k=1}^l Q_{ik}^3} \quad ; \text{ indipendenza } 0 \leq \beta_Y \leq 1 \text{ perfetta dipendenza } y = g(xz) \\ \beta_Z = \frac{C_0}{1 - \sum_{i=1}^n \sum_{j=1}^m P_{ij}^3} \quad ; \text{ indipendenza } 0 \leq \beta_Z \leq 1 \text{ perfetta dipendenza } z = f(xy) \end{array} \right.$$

Indici di Bonferroni trilaterali

$$\left\{ \begin{array}{l} \beta_0 = \sqrt[3]{\beta_X \beta_Y \beta_Z} \quad ; \text{ indipendenza } 0 \leq \beta_0 \leq 1 \text{ perfetta dipendenza trilaterale} \\ \beta_{-1} = \frac{3\beta_X \beta_Y \beta_Z}{\beta_X \beta_Y + \beta_X \beta_Z + \beta_Y \beta_Z} \quad ; \text{ indipendenza } 0 \leq \beta_{-1} \leq 1 \text{ perfetta dipendenza trilaterale} \end{array} \right.$$

essendo: le frequenze marginali:
$$p_i = \sum_{j=1}^m \sum_{k=1}^l f_{ijk} \quad \forall i$$

$$q_j = \sum_{i=1}^n \sum_{k=1}^l f_{ijk} \quad \forall j$$

$$r_k = \sum_{i=1}^n \sum_{j=1}^m f_{ijk} \quad \forall k$$

23 Le espressioni, riportate nel prosieguo, non sono particolarmente complesse, fatto salvo riconoscere come i quadrati delle espressioni bidimensionali siano stati sostituiti dai cubi nelle espressioni tridimensionali, positivi perché riferiti a frequenze e/o probabilità, ed invece rappresentanti geometricamente volumi nello spazio 3D, anziché aree nel piano.

e le frequenze marginali doppie:

$$P_{ij} = \sum_{k=1}^l f_{ijk} \quad \forall ij$$

$$Q_{ik} = \sum_{j=1}^m f_{ijk} \quad \forall ik$$

$$R_{jk} = \sum_{i=1}^n f_{ijk} \quad \forall jk$$

Allora date le due tabelle sotto riportate, di presene e contingenze, le cui dimensioni sono costituite da un numero di strati pari a 4, blocchi pari a 2 e trattamenti pari a 3, gli indici multipli di Bonferroni monolaterali (per righe, colonne e pile) e trilaterali (geometrico ed armonico) risultano:

$$0.511 \quad 0.511 \quad 0.514 \quad 0.512 \quad 0.512$$

	A			B			A			B		
	1	2	3	1	2	3	1	2	3	1	2	3
I	2	2	0	2	0	0	0.042	0.042	-0.042	0.042	-0.042	-0.042
II	2	2	0	2	0	0	0.042	0.042	-0.042	0.042	-0.042	-0.042
III	0	0	2	0	2	2	-0.042	-0.042	0.042	-0.042	0.042	0.042
IV	0	0	2	0	2	2	-0.042	-0.042	0.042	-0.042	0.042	0.042

Questi valori sono abbastanza elevati, anche se non elevatissimi, e danno comunque ragione per continuare uno studio puntuale del fenomeno in esame, mediante l'analisi di varianza.

E.2 Analisi di varianza a tre vie senza interazioni fra le celle

Un insieme di dati stratificato due volte è raggruppabile in celle tridimensionali. L'informazione dentro le celle è rappresentata da una variabile statistica multidimensionale i cui elementi sono i valori degli attributi dell'informazione stessa. Nella tecnica di analisi di varianza si considera la varianza generale come somma di due componenti, le varianze spiegate e la varianza residua, ortogonali tra loro. Lo scopo è quello di massimizzare le varianze spiegate (che descrivono il fenomeno secondo il modello stratificato) rispetto a quella residua (che è legata, invece, al carattere statistico dei dati e, in conclusione, alla loro accidentalità). Nell'analisi di varianza a tre vie senza interazione tra le celle, il modello funzionale adottato è il seguente, mentre il modello stocastico prevede osservazioni indipendenti e di uguale precisione:

$$\bar{a} + \hat{a}_i + \hat{a}_j + \hat{a}_k = s_{ijkl}^o + \hat{v}_{ijkl} \quad \forall i = 1, \dots, I$$

$$\forall j = 1, \dots, J$$

$$\forall k = 1, \dots, K$$

$$\forall l = 1, \dots, L = 1$$

Nella espressione precedente I è il numero trattamenti, J è il numero di blocchi, K è il numero di strati ed $L=1$ è la dimensione della variabile statistica. Il numero complessivo di osservazioni (si veda, a riguardo, la tabella riportata più oltre assieme agli scarti – residui) è $m = I \cdot J \cdot K \cdot L = 24$, ed il numero di parametri e $n = I + J + K + 1 = 10$. A tali equazioni è necessario aggiungere alcune condizioni di vincolo, a causa della

sovraparametrizzazione dovuta ad incognite non linearmente indipendenti che determinano singolarità del sistema. Pertanto per ogni gruppo di parametri, si è imposto che la loro somma fosse esattamente nulla, ovvero una matrice di coefficienti tutti unitari a basso peso (10^{-2}) è stata addizionata alla matrice normale per ogni gruppo di parametri. I valori numerici ad esse relativi possono essere facilmente ottenuti risolvendo, a minimi quadrati, i suddetti sistemi di equazioni d'osservazione e, tramite essi, propagando le varianze delle osservazioni.

SOLUZIONE, SQM E TEST T DI STUDENT

MEDIA:				1
1	0.7426	1.3580	0.5468	2.1100
 STRATI:				4
2	-0.0578	0.5265	-0.1097	2.1100
3	-0.0370	0.5265	-0.0703	2.1100
4	0.0133	0.5265	0.0253	2.1100
5	0.0815	0.5265	0.1547	2.1100
 BLOCCHI:				2
6	0.5025	1.0436	0.4815	2.1100
7	-0.5025	1.0436	-0.4815	2.1100
 TRATTAMENTI:				3
8	-0.2648	0.6977	-0.3796	2.1100
9	0.1123	0.6977	0.1609	2.1100
10	0.1525	0.6977	0.2186	2.1100

Dalla tabella si evidenzia ovviamente la somma nulla dei parametri di ciascun gruppo, ma soprattutto la non significatività di alcun parametro, compresa la media generale, ad eccezione di quello relativo al terzo trattamento, risultato debolmente significativo. Eppure come specificato nel prosieguo, l'analisi di varianza a tre vie è del tutto significativa. Infatti dalle stime dei parametri si ricavano, immediatamente, le varianze spiegate, relative ai trattamenti T , ai blocchi B ed agli strati S , concordemente al modello di riferimento prescelto, e la varianza residua:

$$\sigma_T^2 = \frac{1}{I-1} \sum_{i=1}^I \hat{a}_i^2 \qquad \sigma_B^2 = \frac{1}{J-1} \sum_{j=1}^J \hat{a}_j^2 \qquad \sigma_S^2 = \frac{1}{K-1} \sum_{k=1}^K \hat{a}_k^2$$

$$\sigma_R^2 = \frac{1}{\nu} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L \hat{v}_{ijkl}^2$$

dove il denominatore della varianza residua rappresenta i gradi di libertà del sistema: $\nu = I \cdot J \cdot K \cdot L - (I-1) - (J-1) - (K-1) - 1$. Un giudizio sull'analisi di varianza e sulla validità dei modelli adottati discende dalla comparazione fra la varianza spiegata e residua. In questa sede, si utilizza sempre lo schema rigido di comparazione, confrontando direttamente la varianza spiegata e la varianza residua. Di conseguenza, si ha:

ANALISI DI VARIANZA

STRATI E TEST F DI FISHER	4	0.8126	2.9600
BLOCCHI E TEST F DI FISHER	2	0.0662	3.5900
TRATTAMENTI E TEST F DI FISHER	3	5.8052	3.2000

dove il test F di Fisher passa per quanto riguarda i trattamenti, ovviamente significativi che caratterizzano la prova in esame (e danno senso e significato all'intero trattamento statistico dei dati), mentre lo stesso test non passa per quanto riguarda sia gli strati che i blocchi. Eppure è possibile rilevare come gli strati diano un po' più di significato al test, rispetto ai blocchi, in quanto cambiare strato vuol dire operare ad un diverso livello di profondità, mentre cambiare blocco vuol dire cambiare materiale che le prove simulate a fatica vogliono dimostrare sostanzialmente equivalenti e sostituibili, di conseguenza, essendo il secondo più conveniente, da diversi punti di vista operativi.

Queste considerazioni danno ragione della effettiva validità dell'analisi di varianza, rispetto al semplice test di significatività dei parametri, effettuato con la distribuzione t di Student. In effetti, confrontando qualitativamente la varianza a priori con il sigma zero a posteriori si può osservare una significativa riduzione della dispersione (a riguardo, si noti che, data la loro stretta dipendenza, potrebbero essere confrontati solo in termini non – parametrici) ed il test chi quadrato passa con un'abbondanza addirittura sbalorditiva:

MEDIA, SQM E

N. DATI E PARAMETRI	1.1855	0.3908	20	10
GDL, SIGMA ZERO E				
TEST CHI QUADRATO	17	0.2085	30.2000	

Controlli numerici danno ragione della sostanziale correttezza del modo di procedere. Infatti tanto il numero di condizione del sistema (calcolato sulla norma dell'estremo superiore delle matrici normale ed inversa), quanto le ridondanze locali delle singole osservazioni (calcolate in base al teorema di decomposizione ortogonale della varianza) sono relativamente elevati:

NUMERO DI CONDIZIONE	0.6369
RIDONDANZE LOCALI	0.7083

cosa che ben garantisce di una sufficiente lontananza tanto dalla condizione di singolarità e mal – condizionamento del sistema, quanto dalla condizione di inaffidabilità e non – controllabilità delle singole osservazioni. Da ultimo per quanto riguarda l'eventuale presenza di dati anomali, occorre segnalare come il test tau di Thompson non evidenzi alcuna osservazione, su ventiquattro, oltre la soglia di criticità, cosa che assicura della sufficiente bontà dei dati ed esclude particolari necessità di eliminare dati anomali, prima di effettuare il calcolo di statistiche ottimali.

Nell'immediato prosieguo, si riportano le stime delle osservazioni compatibili con il modello funzionale adottato, precedute dai valori delle osservazioni e seguite dal loro scarto quadratico medio, nonché gli scarti – residui delle equazioni d'osservazione.

DATI, STIME, SQM, SCARTI, SQM, TEST TAU DI THOMPSON

1	0.6828	0.9224	0.1126	0.2396	0.1755	1.3653
2	0.7032	0.9432	0.1126	0.2400	0.1755	1.3672

3	0.7466	0.9935	0.1126	0.2469	0.1755	1.4068
4	0.8056	1.0617	0.1126	0.2561	0.1755	1.4589
5	0.4522	1.2996	0.1126	-0.1526	0.1755	-0.8697
6	1.4639	1.3203	0.1126	-0.1436	0.1755	-0.8182
7	1.5347	1.3706	0.1126	-0.1641	0.1755	-0.9347
8	1.5503	1.4388	0.1126	-0.1115	0.1755	-0.6354
9	1.4522	1.3398	0.1126	-0.1124	0.1755	-0.6403
10	1.4639	1.3606	0.1126	-0.1033	0.1755	-0.5888
11	1.5347	1.4109	0.1126	-0.1238	0.1755	-0.7053
12	1.5503	1.4790	0.1126	-0.0713	0.1755	-0.4060
13	0.1650	-0.0825	0.1126	-0.2475	0.1755	-1.4100
14	0.1902	-0.0617	0.1126	-0.2519	0.1755	-1.4355
15	0.2176	-0.0114	0.1126	-0.2290	0.1755	-1.3047
16	0.3109	0.0567	0.1126	-0.2542	0.1755	-1.4481
17	0.1500	0.2946	0.1126	0.1446	0.1755	0.8241
18	0.1732	0.3154	0.1126	0.1422	0.1755	0.8101
19	0.2203	0.3657	0.1126	0.1454	0.1755	0.8286
20	0.2943	0.4339	0.1126	0.1396	0.1755	0.7952
21	0.2067	0.3349	0.1126	0.1282	0.1755	0.7305
22	0.2389	0.3556	0.1126	0.1167	0.1755	0.6651
23	0.2815	0.4060	0.1126	0.1245	0.1755	0.7093
24	0.4328	0.4741	0.1126	0.0413	0.1755	0.2355

Si noti altresì come, essendo tutti uguali gli scarti quadratici medi delle stime suddette, il teorema di decomposizione ortogonale della varianza faccia sì che anche lo scarto quadratico medio di tutti gli scarti residui sia sempre uguale e pari a 0.7083. Infine giova ribadire come i test di validazione dei modelli, quali quelli qui presentati e svolti, permettano di sottoporre a verifica, mediante opportuni controlli e confronti d'ipotesi, le stime effettuate come, del resto, tutti i risultati ottenuti nell'ambito della statistica.

E.3 Analisi fattoriale

L'analisi fattoriale completa lo studio, in questo caso specifico, iniziato con l'analisi di varianza, effettuando la decomposizione spettrale (cioè separando autovalori ed autovettori) del sistema normale corrispondente, così da stabilire numeri minimi di parametri che sono capaci di spiegare il fenomeno con una certa precisione. Lo stesso studio può altresì essere compiuto mediante la decomposizione ai valori singolari del sistema di equazioni d'osservazione equipesate, tuttavia questa via, se non necessaria per problemi di stabilità numerica, è computazionalmente più onerosa.

Si osservi, a riguardo, come l'equipesatura delle equazioni d'osservazione sia banale, se il modello stocastico è costituito dai pesi delle equazioni, mentre richieda la radice pari della matrice dei cofattori (ovvero della matrice di covarianza delle osservazioni, a meno del fattore di proporzionalità la cui radice quadrata è solitamente detta sigma zero), se sono presenti correlazioni tra le osservazioni. In ogni caso, come evidente nell'immediato prosieguo, i valori singolari della matrice disegno equipesata A_p sono la radice quadrata degli autovalori della matrice normale C :

$$C = R^T AR$$

$$A_p = QST$$

dove A è una matrice diagonale, contenente gli autovalori del sistema normale, ed R una matrice quadrata ortonormale, contenente gli autovettori dello stesso sistema, mentre S è una matrice rettangolare diagonale (alta) con la stessa forma della matrice disegno, contenente i valori singolari del sistema di equazioni d'osservazione, essendo Q e T due matrici quadrate ortonormali di forma prodotto compatibile con la matrice disegno, contenenti gli autovettori del suddetto sistema 24.

A partire dalla matrice normale, il calcolo degli autovalori può essere molto semplicemente effettuato eseguendo, nell'ordine, una sequenza di due operazioni matriciali, costituite da:

- ❑ una fattorizzazione di Cholesky;
- ❑ un prodotto matriciale dei due fattori di Cholesky commutati,

come di seguito riportato:

$$C_1 = C$$

$$T_i : T_i^T T_i = C_i \quad i = 1, 2, \dots$$

$$C_{i+1} = T_i T_i^T$$

Ad ogni passo, la matrice risultante tende a farsi diagonale, facendo via, via diminuire gli elementi extra – diagonali. Il procedimento iterativo è arrestato, quando l'elemento extra – diagonale più grande è inferiore, in valore assoluto, ad un'opportuna soglia prefissata. Ad es., in questo caso specifico, dato un sistema di 9 equazioni in 9 incognite, derivato da quello scritto per l'analisi di varianza, eliminando l'incognita della media generale ed aggiungendo un numero piccolissimo (pari a 10^{-5} ed uguale alla soglia prefissata) a tutti gli elementi diagonali principali, come condizione di regolarizzazione della soluzione cercata, il procedimento iterativo si è arrestato dopo solo 25 iterazioni con un residuo di 1.588D-05. Si riportano, di seguito, gli autovalori trovati ed i loro valori frazionari, rapportati all'unità riferita alla somma.

AUTOVALORI E VAL. FRAZIONARI

1	1.0D-05	0.000	6	8.0D+00	1.111
2	1.0D-05	0.000	7	8.0D+00	1.111
3	6.0D+00	0.833	8	1.2D+01	1.667
4	6.0D+00	0.833	9	2.6D+01	3.611
5	6.0D+00	0.833			

Un primo commento prende atto che, tante volte, sono d'interesse solo gli autovalori e non gli autovettori, come spesso solo le varianze e non le componenti principali, ed allora la strada percorsa, davvero così semplice, sembra davvero molto promettente. Un secondo commento fa riferimento alle dimensioni dei problemi, oggetto di studio, e rileva come l'analisi fattoriale debba essere doverosamente riservata alla selezione di piccoli gruppi di parametri, eventualmente di sotto- problemi, dove ha sicuro interesse.

24 Si osservi inoltre come si abbia: $T=R$ e $Q=A_p T^T S$, essendo S l'inversa generalizzata di S , costruita con gli inversi dei valori singolari e ponendo semplicemente elementi nulli, dove i valori singolari sono nulli, perché lo zero è inverso generalizzato di se stesso.

D'altra parte, non avrebbe senso domandarsi se serve un parametro in più o in meno nella compensazione di una struttura reticolare con decine di migliaia di incognite o nella modellazione digitale di una superficie con centinaia di migliaia di incognite, oppure addirittura nel filtraggio di un'immagine con milioni di incognite. Infatti in tutti questi casi, le varie incognite hanno anche un significato geometrico e/o fisico specifico (ad es., come in un blocco con un gran numero di immagini, oppure in una lunga sequenza sempre di immagini), mentre dove la spiegazione è più vaga ed incerta, allora ha senso mettere in gioco alcuni parametri e domandarsi, quale è il numero minimo di parametri per raggiungere una certa spiegazione.

Il calcolo degli autovettori è un po' più complesso e può essere effettuato mediante l'esecuzione ripetuta di tante rotazioni elementari di Jacobi, fino ad ottenere nuovamente la matrice diagonale degli autovalori ed inoltre la matrice degli autovettori, come prodotto destro di tutte le rotazioni elementari di Jacobi (essendo il prodotto sinistro, con la rotazione elementare di Jacobi inversa, la matrice trasposta della matrice degli autovettori trovata). Esso si articola in tre passaggi:

- selezione dell'elemento extra – diagonale massimo, da annullare:

$$\begin{aligned}
 & -a \sin \alpha \cos \alpha + b \cos^2 \alpha - b \sin^2 \alpha + c \sin \alpha \cos \alpha = 0 \\
 & 2 \cos 2\alpha + \sin 2\alpha = 0 \\
 & \alpha = \frac{1}{2} \arctan \frac{-2b}{c-a}
 \end{aligned}$$

essendo b l'elemento extra – diagonale da annullare
 ed a, c i corrispondenti elementi diagonali principali;

- applicazione della rotazione elementare di Jacobi alla matrice data e via, via trasformata:

detti (h, h') gli indici dell'elemento da annullare ($c_{hh'} = 0$),
 tutti gli elementi restano invariati, tranne quelli appartenenti
 alle righe ed alle colonne h, h' che sono così modificati:

$$\begin{aligned}
 c_{hh} &= \cos^2 \alpha c_{hh} + \sin^2 \alpha c_{h'h'} + 2 \sin \alpha \cos \alpha c_{hh'} \\
 c_{h'h'} &= \sin^2 \alpha c_{hh} + \cos^2 \alpha c_{h'h'} - 2 \sin \alpha \cos \alpha c_{hh'} \\
 c_{ih} &= \cos \alpha c_{ih} + \sin \alpha c_{ih'} \quad \forall i \neq h, h' \\
 c_{ih'} &= -\sin \alpha c_{ih} + \cos \alpha c_{ih'} \\
 c_{hj} &= \cos \alpha c_{hj} + \sin \alpha c_{h'j} \quad \forall j \neq h, h' \\
 c_{h'j} &= -\sin \alpha c_{hj} + \cos \alpha c_{h'j}
 \end{aligned}$$

dove qui, come al passo precedente, è relativamente facile ottenere
 tutte le espressioni cercate, facendo uso del prodotto di Kronecker
 che permette di eseguire velocemente il prodotto di tre matrici,

- prodotto destro delle rotazioni elementari trovate, rimanendo invariati tutti gli altri elementi:

$$\begin{aligned}
 r_{ih} &= \cos \alpha r_{ih} + \sin \alpha r_{ih'} \\
 r_{ih'} &= -\sin \alpha r_{ih} + \cos \alpha r_{ih'}
 \end{aligned}$$

Ad ogni passo, la matrice risultante annulla l'elemento selezionato, ovvero il massimo elemento extra – diagonale 25, ancora una volta, facendo diminuire tutti gli elementi extra – diagonali. Il procedimento iterativo è arrestato, quando l'elemento extra – diagonale più grande è inferiore, in valore assoluto, ad un'opportuna soglia prefissata. Ad es., in questo caso specifico, dato lo stesso sistema di 9 equazioni in 9 incognite, il procedimento iterativo si è arrestato dopo ben 91 iterazioni con un residuo di $2.424D-10$, avendo dovuto fissare come soglia d'arresto del ciclo iterativo un numero ancora più piccolo, pari a 10^{-10} , pena la non convergenza del procedimento iterativo. Si riporta, di seguito, la matrice degli autovettori trovata.

MATRICE DEGLI AUTOVETTORI

	1	2	3	4	5	6	7	8	9
1	0.454	0.158	0.000	0.000	0.000	0.413	-0.705	0.000	0.320
2	0.454	0.158	0.000	0.000	0.000	0.404	0.710	0.000	0.320
3	0.454	0.158	0.000	0.000	0.000	-0.816	-0.005	0.000	0.320
4	-0.437	0.280	0.000	0.000	0.000	0.000	0.000	0.707	0.480
5	-0.437	0.280	0.000	0.000	0.000	0.000	0.000	-0.707	0.480
6	-0.017	-0.438	0.741	-0.370	-0.253	0.000	0.000	0.000	0.240
7	-0.017	-0.438	-0.670	-0.453	-0.310	0.000	0.000	0.000	0.240
8	-0.017	-0.438	-0.036	0.811	-0.302	0.000	0.000	0.000	0.240
9	-0.017	-0.438	-0.035	0.012	0.865	0.000	0.000	0.000	0.240

Il calcolo delle componenti principali prevede la soluzione di più sistemi, a partire da quello con tutti gli autovalori trovati (che ovviamente dà proprio la stessa soluzione a minimi quadrati), per proseguire annullando progressivamente alcuni autovalori, incominciando dai più piccoli, così da vedere quanto si degrada via, via la soluzione stessa. Si noti, a riguardo, come l'inverso generalizzato, di minima norma, di un autovalore nullo sia nullo, a sua volta. Allora con specifico riferimento al caso in esame, di seguito, si riportano le soluzioni alle componenti principali, senza togliere alcun autovalore, togliendo due autovalori nulli poi altri tre non troppo grandi ed infine altri due un po' più grandicelli. Le soluzioni ottenute mostrano che, in questo caso specifico, anche solo 4 parametri bastano a spiegare circa il 75% della variabilità presente nei dati, senza sostanziali perdite di precisione.

SOLUZIONI ALLE COMPONENTI PRINCIPALI (VALORI ANNULLATI: 2 5 7)
SIGMA ZERO, N. EQUAZIONI E INCOGNITE (RIDONDANZE: 15 17 20 22)
0.222 24 9 0.222 24 7 0.232 24 4 0.332 24 2

SOLUZIONI E SQM

1	-0.036	0.062	-0.036	0.062	-0.036	0.059	0.228	0.017
2	0.341	0.062	0.341	0.062	0.341	0.059	0.228	0.017
3	0.381	0.062	0.381	0.062	0.381	0.059	0.228	0.017
4	0.845	0.047	0.845	0.047	0.845	0.045	0.845	0.062
5	-0.160	0.047	-0.160	0.047	-0.160	0.045	-0.160	0.062
6	0.114	0.074	0.114	0.074	0.171	0.009	0.171	0.013
7	0.134	0.074	0.134	0.074	0.171	0.009	0.171	0.013
8	0.185	0.074	0.185	0.074	0.171	0.009	0.171	0.013
9	0.253	0.074	0.253	0.074	0.171	0.009	0.171	0.013

25 Si noti come la ricerca del massimo sia accelerata, memorizzando ad ogni passaggio, successivo al primo, il valore massimo di ogni riga. Infatti poiché l'algorithmo modifica, ad ogni passo, solo due righe e due colonne, basta eseguire un confronto a tre per tutte le righe, tranne quelle modificate, e due confronti completi, limitatamente alle due righe modificate, oltreché cercare il massimo dei massimi.

TEORIA DELLA PROBABILITÀ

Definizione di probabilità

Probabilità geometrica o di Laplace: eventi indistinguibili → uguale probabilità (limitata alla Teoria dei Giochi).
Probabilità frequentista o di Von Mises: limite delle frequenze (empiriche) (vera solo in probabilità e pertanto contraddittoria).
Probabilità assiomatica o di Kolmogorov: $0 \leq P_A \leq 1$
 $P_A = 0$ se $A \equiv \emptyset$
 $[P_A = 1$ se $A \equiv S$ (universo)]
 $P_A \geq P_B$ se $A \supseteq B$
 $P_{A+B} = P_A + P_B$ se $AB \equiv \emptyset$ (eventi incompatibili).

Teoremi della probabilità

Totale: $P_{A+B} = P_A + P_B - P_{AB}$ anche se $AB \neq \emptyset$ (eventi compatibili).
Composta: $P_{AB} = P_A P_B$ solo se A, B (eventi indipendenti),
 $P_{AB} = P_A P_{B/A} = P_{A/B} P_B$ anche se A, B (eventi dipendenti).
di Bayes: $P_{A/B} = \frac{P_{B/A} P_A}{P_B}$; $P_{B/A} = \frac{P_{A/B} P_B}{P_A}$

STATISTICA DESCRITTIVA

VARIABILE STATISTICA SEMPLICE

- Momento semplice di ordine k : $\mu_k = \sum_{i=1}^n x_i^k f_i$ $\mu_k = \frac{1}{N} \sum_{i=1}^n x_i^k$ se $f_i = \frac{1}{N}$
- Momento centrale di ordine k : $\bar{\mu}_k = \sum_{i=1}^n (x_i - \mu)^k f_i$ $\bar{\mu}_k = \frac{1}{N} \sum_{i=1}^n (x_i - \mu)^k$ se $f_i = \frac{1}{N}$
- Relazioni che legano i momenti centrali ai momenti semplici:

$$\bar{\mu}_2 = \mu_2 - \mu^2 = \sigma^2 \quad ; \quad \bar{\mu}_2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \mu^2 = \sigma^2$$

$$\bar{\mu}_3 = \mu_3 - 3\mu_2\mu + 2\mu^3 \quad ; \quad \bar{\mu}_3 = \frac{1}{N} \sum_{i=1}^N x_i^3 - \frac{3}{N} \mu \sum_{i=1}^N x_i^2 + 2\mu^3$$

$$\bar{\mu}_4 = \mu_4 - 4\mu_3\mu + 6\mu_2\mu^2 - 3\mu^4 \quad ; \quad \bar{\mu}_4 = \frac{1}{N} \sum_{i=1}^N x_i^4 - \frac{4}{N} \mu \sum_{i=1}^N x_i^3 + \frac{6}{N} \mu^2 \sum_{i=1}^N x_i^2 - 3\mu^4$$

²⁶ Questo sommario riprende, estende e generalizza quanto raccolto e presentato da Federica Migliaccio, con un'accurata ed approfondita indagine metodologica, avendo avuto ella stessa la titolarità dell'insegnamento di Statistica per il Corso di Laurea in PTUA della Facoltà di Architettura del Politecnico di Milano; a lei va il ringraziamento dell'autore.

Resta da osservare come questo sommario voglia appositamente essere un sommario, collocato come l'ultima appendice di un lavoro di statistica sui modelli ottimali per l'analisi dei dati che ha, tra i suoi primi paragrafi, un'illustrazione discorsiva, ma esauriente della statistica (dalla statistica descrittiva all'inferenza statistica, passando per la teoria della stima). Pertanto questa appendice / sommario non è un capitolo di un libro, oltretutto fuori posto, rispetto alla tradizionale collocazione, ma un utile strumento di ripasso e/o richiamo, per la lettura dello stesso lavoro, dove è collocata/o, come pure di altri lavori vicini per i quali può essere ugualmente utile. Infine per quanto riguarda eventuali rimandi a libri di statistica elementare, corre il dovere di dire questa è pressoché sterminata e notevole anche in lingua italiana, andando da libri antichi e classici, veri e propri testi base per uno studio accurato, a testi moderni ed innovativi, capaci di prendere in seria considerazione anche le moderne tecnologie dell'informazione e della comunicazione.

Nota: N = numerosità dei dati;
 n = numerosità degli argomenti della variabile X (ovvero numero di classi).
 $n = 1 + 3.322 \log n$ regola di Sturges (solitamente comunque: $5 \leq n \leq 25$)

Indici di posizione (o di centro)

I centri di ordine r minimizzano la funzione: $z = \sum_{i=1}^n |x_i - c_r|^r f_i$

Media: $c_2 = \mu = \sum_{i=1}^n x_i f_i$ $r = 2$
Mediana: $c_1 = \mu_e$ $r = 1$ *valore che divide in due parti uguali l'istogramma*
Moda: $c_0 = \mu_k$ $r = 0$ *valore con la frequenza più alta*

Indici di dispersione

Varianza (la sua radice quadrata è detta s.q.m.): $\sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 f_i$ $r = 2$

M.A.V. (deviazione media assoluta dalla mediana): $\sigma_e = \sum_{i=1}^n |x_i - \mu_e| f_i$ $r = 1$

m.a.v. (mediana dei moduli rispetto alla mediana): $\delta_e = \text{mediana di } |x_i - \mu_e|$

Ampiezza: $\Delta = x_{max} - x_{min}$

Indici di forma

Asimmetria: $\gamma = \frac{\sum_{i=1}^n (x_i - \mu)^3 f_i}{\sigma^3}$; $\Gamma = \frac{c_1 - c_0}{\delta_e}$ $\gamma; \Gamma < 0$ *coda a sinistra*
 $\gamma; \Gamma = 0$ *simmetrica*
 $\gamma; \Gamma > 0$ *coda a destra*

Curtosi: $\beta = \frac{\sum_{i=1}^n (x_i - \mu)^4 f_i}{\sigma^4}$ $\beta > 3$ *leptocurtica*
 $\beta = 3$ *normale*
 $\beta < 3$ *platicurtica*

Nota: La media, la varianza ed il coefficiente di correlazione (si veda oltre) sono indici ottimali; la mediana, il m.a.v. e gli indici di Bonferroni (si veda sempre oltre) sono indici robusti.

Teorema di Chebychev $P(X) \geq 1 - \frac{1}{\lambda^2}$ *Probabilità normali*

$\lambda = 1$	$P(X) \geq 0$	$P_N(X) = 0.68 \cong 0.70$
$\lambda = 2$	$P(X) \geq 0.75$	$P_N(X) = 0.95$
$\lambda = 3$	$P(X) \geq 0.89 \cong 0.90$	$P_N(X) = 0.997$
...		
$\lambda = 5$	$P(X) \geq 0.96 \cong 0.95$	
$\lambda = 10$	$P(X) \geq 0.99$	
$\lambda = 18 \cong 20$	$P(X) \geq 0.997$	

Nota: Il Teorema di Chebychev istituisce un primo confronto fra i risultati degli esperimenti: le variabili statistiche, ed i modelli interpretativi: le variabili casuali, al di là della loro nota identità formale.

Nota sulle medie

❑ Radice della media quadratica: $x_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$

❑ Media geometrica: $x_0 = \sqrt[n]{\prod_{i=1}^n x_i}$

❑ Media armonica: $x_{-1} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$

❑ Media ponderata (potata): $\bar{x}_p = \frac{\sum_{i=1}^n p_i x_i}{\sum_{i=1}^n p_i}$

$p_i =$ peso (se: $p_i = 0$, per qualche i)

e sulla mutua variabilità

- Delta di Gini:

$$\Delta = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{n \cdot (n-1)} \quad (\forall j \neq i)$$

- Rapporto di concentrazione:

$$nulla \quad 0 \leq R = \frac{\Delta}{2\mu} \leq 1 \quad max$$

- Entropia di Shannon:

$$H = -\sum_{i=1}^n f_i \log_c f_i$$

$$nulla \quad 0 \leq \frac{H}{\log_c n} \leq 1 \quad max$$

- $c = 2$ (teoria dell'informazione)
- $c = e$ (variabili casuali continue)

VARIABILE STATISTICA DOPPIA

Y X	y_1	y_2	...	y_j	...	y_m	p_i
x_i	f_{11}	f_{12}	...	f_{1j}	...	f_{1m}	p_1
x_i	f_{21}	f_{22}	...	f_{2j}	...	f_{2m}	p_2
\vdots	\vdots					\vdots	\vdots
x_i	f_{i1}	f_{i2}	...	f_{ij}	...	f_{im}	p_i
\vdots	\vdots					\vdots	\vdots
x_i	f_{n1}	f_{n2}	...	f_{nj}	...	f_{nm}	p_n
q_j	q_1	q_2	...	q_j	...	q_m	1

Nota:

Le variabili statistiche sono sempre risultati di esperimenti e, di conseguenza, concrete, finite e discrete.

Le variabili casuali sono modelli interpretativi e, in generale, astratte, illimitate e continue. Un'eccezione è costituita dalla variabili casuali della Teoria dei Giochi che sono sempre astratte, ma finite e discrete.

$f_{ij} =$ frequenze relative $\sum_{i=1}^n \sum_{j=1}^m f_{ij} = 1$;

$N =$ numerosità dei dati,

$p_i =$ frequenze marginali X $\sum_{j=1}^m f_{ij} = p_i$;

$m =$ numerosità degli argomenti della variabile X ,

$q_j =$ frequenze marginali Y $\sum_{i=1}^n f_{ij} = q_j$;

$n =$ numerosità degli argomenti della variabile Y .

Connessione (utile per basse dipendenze)

Nota:

Contingenze: $c_{ij} = f_{ij} - p_i q_j$
 $-1 \leq c_{ij} \leq 1$

Le variabili casuali normali hanno quali momenti caratteristici solo media/e e varianza/e; se a più dimensioni, l'unica dipendenza ammessa è quella lineare, espressa da coefficiente/i di correlazione lineare semplice che, insieme alla/e varianza/e formano la matrice di varianza – covarianza.

Semicontingenza media: $C_0 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m |c_{ij}|$

Indici di Bonferroni unilaterali:

$$\left\{ \begin{array}{l} \beta_X = \frac{c_0}{1 - \sum_{i=1}^n p_i^2} \quad ; \text{ indipendenza } 0 \leq \beta_X \leq 1 \text{ perfetta dipendenza } x = h(y) \\ \beta_Y = \frac{c_0}{1 - \sum_{j=1}^m q_j^2} \quad ; \text{ indipendenza } 0 \leq \beta_Y \leq 1 \text{ perfetta dipendenza } y = g(x) \end{array} \right.$$

Indici di Bonferroni bilaterali:

$$\left\{ \begin{array}{l} \beta_0 = \sqrt{\beta_X \beta_Y} \quad ; \text{ indipendenza } 0 \leq \beta_0 \leq 1 \text{ perfetta dipendenza bilaterale} \\ \beta_{-1} = \frac{2\beta_X \beta_Y}{\beta_X + \beta_Y} \quad ; \text{ indipendenza } 0 \leq \beta_{-1} \leq 1 \text{ perfetta dipendenza bilaterale} \end{array} \right.$$

Correlazione (dipendenza lineare) Distribuzioni marginali:

$$X \left\{ \begin{array}{l} x_1 \dots x_i \dots x_n \\ p_1 \dots p_i \dots p_n \end{array} \right. \quad Y \left\{ \begin{array}{l} y_1 \dots y_i \dots y_n \\ q_1 \dots q_i \dots q_n \end{array} \right.$$

Media della variabile marginale X: $\mu_X = \sum_{i=1}^n x_i p_i$

Media della variabile marginale Y: $\mu_Y = \sum_{j=1}^m y_j q_j$

Varianza della variabile marginale X: $\sigma_X^2 = \sum_{i=1}^n (x_i - \mu_X)^2 p_i$

Varianza della variabile marginale Y: $\sigma_Y^2 = \sum_{j=1}^m (y_j - \mu_Y)^2 q_j$

Covarianza tra le variabili X e Y: $\sigma_{XY} = \sum_{i=1}^n \sum_{j=1}^m (x_i - \mu_X)(y_j - \mu_Y) f_{ij} = \sum_{i=1}^n \sum_{j=1}^m x_i y_j f_{ij} - \mu_X \mu_Y$

Coefficiente di correlazione lineare:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad ; \text{ perfetta dipendenza inversa } -1 \leq \rho \leq 1 \text{ perfetta dipendenza diretta}$$

$$\rho = 0 \text{ indipendenza lineare}$$

Rette di regressione:

$$Y = aX + b$$

$$X = cY + d$$

$$a = \frac{\sigma_{XY}}{\sigma_X^2}$$

$$c = \frac{\sigma_{XY}}{\sigma_Y^2}$$

$$b = \mu_Y - a\mu_X$$

$$d = \mu_X - c\mu_Y$$

Rette di regressione robuste:

$$Y = aX + b$$

$$a = \text{mediana}(a_{ij}) \quad \forall i, j > i$$

$$b = \text{mediana}(y) - a \text{mediana}(x)$$

$$X = cY + d$$

$$c = \text{mediana}(c_{ij}) \quad \forall i, j > i$$

$$d = \text{mediana}(x) - c \text{mediana}(y)$$

Regressione (utile per alte dipendenze)

Distribuzioni condizionate:

$$X|_{y_j} \begin{cases} x_1 & \dots & x_i & \dots & x_n \\ \frac{f_{1j}}{q_j} & \dots & \frac{f_{ij}}{q_j} & \dots & \frac{f_{nj}}{q_j} \end{cases} \quad Y|_{x_i} \begin{cases} y_1 & \dots & y_i & \dots & y_n \\ \frac{f_{i1}}{p_i} & \dots & \frac{f_{ij}}{p_i} & \dots & \frac{f_{in}}{p_i} \end{cases}$$

Media di X condizionata a $y = y_j$:
$$\mu_{x|y_j} = \sum_{i=1}^n \frac{x_i f_{ij}}{q_j} \quad (i = 1, \dots, n)$$

Media di Y condizionata a $x = x_i$:
$$\mu_{y|x_i} = \sum_{j=1}^m \frac{y_j f_{ij}}{p_i} \quad (j = 1, \dots, m)$$

Varianza di X condizionata a $y = y_j$:
$$\sigma_{x|y_j}^2 = \sum_{i=1}^n \frac{(x_i - \mu_{x|y_j})^2 f_{ij}}{q_j} \quad (i = 1, \dots, n)$$

Varianza di Y condizionata a $x = x_i$:
$$\sigma_{y|x_i}^2 = \sum_{j=1}^m \frac{(y_j - \mu_{y|x_i})^2 f_{ij}}{p_i} \quad (j = 1, \dots, m)$$

Nota: Le medie delle medie condizionate coincidono con le medie generali.

Varianza delle medie condizionate:
$$\sigma_{S_{x|y}}^2 = \sum_{j=1}^m (\mu_{x|y_j} - \mu_X)^2 q_j$$

Varianza delle medie condizionate:
$$\sigma_{S_{y|x}}^2 = \sum_{i=1}^n (\mu_{y|x_i} - \mu_Y)^2 p_i$$

Media delle varianze condizionate:
$$\sigma_{R_{x|y}}^2 = \sum_{j=1}^m \sigma_{x|y_j}^2 q_j$$

Media delle varianze condizionate:
$$\sigma_{R_{y|x}}^2 = \sum_{i=1}^n \sigma_{y|x_i}^2 p_i$$

Teorema di decomposizione ortogonale della varianza:
$$\sigma_{S_{x|y}}^2 + \sigma_{R_{x|y}}^2 = \sigma_X^2$$

$$\sigma_{S_{y|x}}^2 + \sigma_{R_{y|x}}^2 = \sigma_Y^2$$

Indici di Pearson monolaterali (max indice di dipendenza: $\sigma_S^2 = \max$ se $\sigma_R^2 = \min$):

$$\eta_X^2 = \frac{\sigma_{S_{x|y}}^2}{\sigma_X^2} = 1 - \frac{\sigma_{R_{x|y}}^2}{\sigma_X^2} \quad ; \quad \text{indipendenza in media} \quad 0 \leq \eta_X^2 \leq 1 \quad \text{perfetta dipendenza } x = h(y)$$

$$\eta_Y^2 = \frac{\sigma_{S_{y|x}}^2}{\sigma_Y^2} = 1 - \frac{\sigma_{R_{y|x}}^2}{\sigma_Y^2} \quad ; \quad \text{indipendenza in media} \quad 0 \leq \eta_Y^2 \leq 1 \quad \text{perfetta dipendenza } y = g(x)$$

Indice di Pearson bilaterale (max indice di dipendenza: $\sigma_S^2 = \max$ se $\sigma_R^2 = \min$):

$$\eta^2 = \frac{\sigma_X^2 \eta_X^2 + \sigma_Y^2 \eta_Y^2}{\sigma_X^2 + \sigma_Y^2} \quad ; \quad \text{indipendenza in media} \quad 0 \leq \eta^2 \leq 1 \quad \text{perfetta dipendenza bilaterale}$$

Regressione robusta

Un'alternativa alla regressione classica è data dalla regressione robusta dove le mediane sono al posto delle medie ed i mav al posto delle varianze; in questo caso non vale ovviamente il teorema di decomposizione ortogonale della varianza: comunque il mav spiegato è dato dal mav delle mediane condizionate ed il mav residuo dalla mediana dei mav condizionati.

TRASFORMAZIONE DI VARIABILI CASUALI

In generale, una trasformazione dalla variabile casuale \underline{X} alla variabile casuale \underline{Y} può essere scritta come:

$$\underline{Y} = g(\underline{X})$$

in particolare, una trasformazione non - singolare, dalla v.c. \underline{X} (a n dimensioni) alla v.c. \underline{Y} (a $m \leq n$ dimensioni) costituisce una trasformazione regolare (se $m = n$), oppure altrimenti una contrazione:

$$\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \longrightarrow \underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix}$$

Supponendo note media e matrice di covarianza di \underline{X} :

$$\mu_X = \begin{pmatrix} \mu_{X_1} \\ \mu_{X_2} \\ \vdots \\ \mu_{X_n} \end{pmatrix} ; \quad C_{XX} = \begin{pmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \dots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \dots & \sigma_{X_2 X_n} \\ \vdots & \vdots & \dots & \vdots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \dots & \sigma_{X_n}^2 \end{pmatrix}$$

si vogliono determinare media e covarianza di \underline{Y} .

- Per quanto riguarda il valore medio μ_Y , se \underline{X} è ben concentrato attorno a μ_X e nella stessa zona e la funzione $\underline{Y} = g(\underline{X})$ è gradualmente variata, si può utilizzare l'espressione approssimata:

$$\mu \cong g(\mu_X)$$

Nota: Se $\underline{Y} = g(\underline{X}) = A\underline{X} + b$ è una funzione lineare, l'espressione $\underline{Y} = g(\mu_X) = A\mu_X + b$ è esatta.

- Per quanto riguarda la matrice di covarianza C_{YY} , per semplicità, si distinguono due casi:

1. Caso lineare: la trasformazione da \underline{X} a \underline{Y} è di tipo lineare: $\underline{Y} = A\underline{X} + b$

Pertanto data la matrice dei coefficienti della trasformazione lineare:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

la legge di propagazione della covarianza risulta: $C_{YY} = AC_{XX}A^T$

2. Caso non lineare: la trasformazione da \underline{X} a \underline{Y} è di tipo non lineare: $\underline{Y} = F(\underline{X})$

In questo caso è necessario introdurre lo Jacobiano (riportato appena sotto), cioè la matrice delle derivate parziali delle funzioni $g_i(\underline{X})$ rispetto alle componenti di \underline{X} , ed allora la legge di propagazione della covarianza risulta:

$$C_{YY} = JC_{XX}J^T$$

Matrice Jacobiano:

$$J = \begin{vmatrix} \frac{\partial g_1}{\partial X_1} & \frac{\partial g_1}{\partial X_2} & \dots & \frac{\partial g_1}{\partial X_n} \\ \frac{\partial g_2}{\partial X_1} & \frac{\partial g_2}{\partial X_2} & \dots & \frac{\partial g_2}{\partial X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial X_1} & \frac{\partial g_m}{\partial X_2} & \dots & \frac{\partial g_m}{\partial X_n} \end{vmatrix}$$

Nota: In base alla legge di propagazione della varianza ed alla definizione di varianza dei momenti, la varianza della media e la varianza della varianza di una variabile casuale, rispettivamente qualsiasi e normale ($\beta = 3$), costituita da un insieme di dati indipendenti fra loro, hanno espressione:

$$\sigma_x^2 = \frac{\sigma^2}{n}$$

$$\sigma_{\sigma^2}^2 = \frac{\beta \sigma^4 - \sigma^4}{n} = \frac{2\sigma^4}{n}$$

Nota: $m = 1$; $m = 2$

$$\sigma_Y^2 = \left(\frac{\partial g}{\partial X_1} \right)^2 \sigma_{X_1}^2 + \left(\frac{\partial g}{\partial X_2} \right)^2 \sigma_{X_2}^2 + 2 \frac{\partial g}{\partial X_1} \frac{\partial g}{\partial X_2} \sigma_{X_1 X_2}$$

Esempi:

$$y = x_1 \pm x_2$$

$$\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 \pm 2\sigma_{x_1 x_2}$$

$$y = x_1 x_2$$

$$\sigma_y^2 = \frac{y^2}{x_1^2} \sigma_{x_1}^2 + \frac{y^2}{x_2^2} \sigma_{x_2}^2 + 2 \frac{y^2}{x_1 x_2} \sigma_{x_1 x_2}$$

$$y = \frac{x_1}{x_2}$$

$$\sigma_y^2 = \frac{y^2}{x_1^2} \sigma_{x_1}^2 + \frac{y^2}{x_2^2} \sigma_{x_2}^2 - 2 \frac{y^2}{x_1 x_2} \sigma_{x_1 x_2}$$

In entrambi i casi, indicata con A , sia la matrice dei coefficienti della trasformazione lineare, sia lo Jacobiano, ovvero la matrice delle derivate parziali delle funzioni $g_i(\underline{X})$ rispetto alle componenti di \underline{X} , si possono scrivere le seguenti espressioni algebriche che sono dette leggi di propagazione della varianza e della covarianza, ovviamente del tutto analoghe alle suddette espressioni matriciali:

$$\sigma_{y_i}^2 = \sum_{k=1}^n a_{ik}^2 \sigma_{x_k}^2 + 2 \sum_{k=1}^{n-1} \sum_{l=k+1}^n a_{ik} a_{il} \sigma_{x_k x_l} \quad \forall i$$

$$\sigma_{y_i y_j} = \sum_{k=1}^n a_{ik} a_{jk} \sigma_{x_k}^2 + \sum_{k=1}^{n-1} \sum_{l=k+1}^n (a_{ik} a_{jl} + a_{jk} a_{il}) \sigma_{x_k x_l} \quad \forall i, j \geq 1$$

Più in generale, data una variabile casuale \underline{X} , con funzione densità di probabilità $p(\underline{X})$, la funzione densità di probabilità $p(\underline{Y})$ della variabile casuale \underline{Y} , si ottiene effettuando una trasformazione di distribuzione, grazie al teorema di conservazione delle probabilità elementari, e nel caso regolare in cui le dimensioni delle due variabili casuali sono identiche ($n = m$) risulta:

$$p(\underline{Y}) = \frac{p(\underline{X})}{|\det J|} \quad \text{Nota:} \quad n = m = 1 \quad p(y) = \frac{p(x)}{|g'(x)|}$$

mentre nel caso di una contrazione ($m \leq n$), limitatamente al caso particolare in cui $m = 1$, $n = 2$ e la trasformazione sia una somma fra due variabili casuali indipendenti, si ha l'integrale di convoluzione:

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

Si osservi infine come il caso di un'espansione ($m \geq n$) dia luogo ad distribuzioni singolari.

ESEMPI DI VARIABILI CASUALI

□ Variabile casuale normale (per gli errori accidentali di misura):

Funzione densità di probabilità e funzione distribuzione di probabilità per il caso monodimensionale:

(variabile qualsiasi)

(variabile standardizzata: $\bar{z} = 0$; $\sigma_z^2 = 1$)

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

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{con} \quad z = \frac{x-\bar{x}}{\sigma}$$

$$P(x) = \text{erf}(x)$$

$$P(z) = \text{erf}(z) = \sigma \text{erf}(x)$$

Nota: Funzione densità di probabilità per il caso bidimensionale:

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\bar{x})^2}{\sigma_x^2} - 2\rho\frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2}\right]}$$

semiassi ed orientamento dell'ellisse d'errore: $\lambda_{max/min} = \frac{\sigma_x^2 + \sigma_y^2}{2} \pm \frac{1}{2}\sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2}$

$$\tan 2\vartheta = \frac{-2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2}$$

□ Variabile casuale rettangolare (per le parti piccole):

Funzione densità di probabilità:

$$p(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

$$p(x) = 0 \quad x \leq a ; x \geq b$$

Trasformazione alla v.c. normale: $z = \text{erf}^{(-1)}(x)$; $p(z) = N(z) p(x)$

□ Variabile casuale log - normale (per gli eventi rari):

Funzione densità di probabilità:

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

Trasformazione alla v.c. normale: $z = \ln x$; $p(z) = x p(x)$

Teoremi della normalità e teoremi limite

La distribuzione normale è caratterizzata da media e varianza, nel caso monodimensionale, da due medie e due varianze, oltreché dal coefficiente di correlazione, nel caso bidimensionale, dai vettori delle medie e delle varianze, nonché dalla matrice dei coefficienti di correlazione, nel caso a n dimensioni. Essa è simmetrica e normocurtica, invariante per trasformazioni lineari e per convoluzione con perturbazioni infinitesime qualsiasi. Per il Teorema di Gauss o "limite centrale della statistica", essa è, in legge, la distribuzione limite delle combinazioni lineari di variabili casuali qualsiasi, purché indipendenti e con varianza comparabile, fatto che dà valore alle operazioni di media e media ponderata delle osservazioni. Si rammenta inoltre che il Teorema di Bernoulli o "legge dei grandi numeri" garantisce, in probabilità, la convergenza delle variabili statistiche alle variabili casuali, fatto che avvalorava il confronto fra i risultati degli esperimenti: le variabili statistiche, ed i modelli interpretativi: le variabili casuali, al di là della loro nota identità formale.

INFERENZA STATISTICA - TEST PARAMETRICI

Livello di significatività α od errore di 1^a specie: rischio di rifiutare un'ipotesi vera.

Nota: l'ipotesi d'indipendenza è sempre indispensabile.

Test sulla media di campioni numerosi:

Intervalli fiduciari:

□ Ipotesi $H_0: \mu = \mu_0$ (σ^2 noto): $\frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{N}}} \approx z$ $\bar{x} - \frac{z\sigma}{\sqrt{N}} \leq \mu_0 \leq \bar{x} + \frac{z\sigma}{\sqrt{N}}$

□ Ipotesi $H_0: \mu = \mu_0$ (σ^2 incognito): $\frac{|\bar{x} - \mu_0|}{\frac{\hat{\sigma}}{\sqrt{N}}} \approx z$ $\bar{x} - \frac{z\hat{\sigma}}{\sqrt{N}} \leq \mu_0 \leq \bar{x} + \frac{z\hat{\sigma}}{\sqrt{N}}$

□ Ipotesi $H_0: \mu_X = \mu_Y$ (σ_X^2, σ_Y^2 noti): $\frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_X^2}{N_X} + \frac{\sigma_Y^2}{N_Y}}} \approx z$

□ Ipotesi $H_0: \mu_X = \mu_Y$ (σ_X^2, σ_Y^2 incogniti): $\frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\hat{\sigma}_X^2}{N_X} + \frac{\hat{\sigma}_Y^2}{N_Y}}} \approx z$

Note: Nel caso in cui si suppone: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, è opportuno dare un'unica stima della stessa varianza e, in tal caso, l'espressione utilizzata per la verifica dell'ipotesi H_0 diventa:

$$\hat{\sigma}^2 = \frac{(N_X - 1) \cdot \hat{\sigma}_X^2 + (N_Y - 1) \cdot \hat{\sigma}_Y^2}{N_X + N_Y - 2} ; \quad \frac{|\bar{x} - \bar{y}|}{\hat{\sigma} \sqrt{\frac{1}{N_X} + \frac{1}{N_Y}}} \approx z$$

Nel caso in cui si suppone anche: $\mu_X = \mu_Y = \mu$, è opportuno dare un'unica stima della stessa media, oltreché della varianza:

$$\bar{\mu} = \frac{N_X \cdot \bar{x} + N_Y \cdot \bar{y}}{N_X + N_Y} \quad \hat{\sigma}^2 = \frac{(N_X - 1) \cdot \hat{\sigma}_X^2 + (N_Y - 1) \cdot \hat{\sigma}_Y^2 + N_X N_Y (\bar{x} - \bar{y})^2}{N_X + N_Y - 1}$$

Curva di potenza della distribuzione normale (potenza del test β od errore di 2^a specie: rischio di accettare un'ipotesi falsa), un esempio:

z	0	1	2	2.5	3	4	5
$1 - \beta$ $\alpha = 5\%$		15%	50%		85%	~100%	
$1 - \beta$ $\alpha = 2\%$		10%	38%		76%	96%	~100%
$1 - \beta$ $\alpha = 1\%$		7%	30%	50%	70%	93%	~100%

Test sui parametri di campioni normali:

Intervalli fiduciari:

□ Ipotesi $H_0: \mu = \mu_0$ (σ^2 incognito): $\frac{|\bar{x} - \mu_0|}{\frac{\hat{\sigma}}{\sqrt{N}}} \approx t_{N-1}$ $\bar{x} - \frac{t_v \hat{\sigma}}{\sqrt{N}} \leq \mu_0 \leq \bar{x} + \frac{t_v \hat{\sigma}}{\sqrt{N}}$

- Ipotesi $H_0: \mu_X = \mu_Y$ (σ_X^2, σ_Y^2 incogniti):

$$\frac{|\bar{x} - \bar{y}|}{\hat{\sigma} \sqrt{\frac{1}{N_X} + \frac{1}{N_Y}}} = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{N_X + N_Y}{N_X N_Y} \cdot \frac{(N_X - 1) \cdot \hat{\sigma}_X^2 + (N_Y - 1) \cdot \hat{\sigma}_Y^2}{N_X + N_Y - 2}}} \approx t_{N_X + N_Y - 2}$$

Nota: E' indispensabile che sia: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, del quale si dà la stima $\hat{\sigma}^2$.

- Ipotesi $H_0: \sigma^2 = \sigma_0^2$: $\frac{\hat{\sigma}^2}{\sigma_0^2} (N - 1) \approx \chi_{N-1}^2$ $\frac{\hat{\sigma}^2}{\chi_{\nu}^{2(+)}} \leq \sigma_0^2 \leq \frac{\hat{\sigma}^2}{\chi_{\nu}^{2(-)}}$

- Ipotesi $H_0: \sigma_X^2 = \sigma_Y^2$: $\frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \approx F_{N_X - 1, N_Y - 1}$

- Ipotesi $H_0: \rho = 0$ (N deve essere sufficientemente grande):

$$\frac{\hat{r}_{XY}}{\sigma_r} = \frac{\hat{r}_{XY}}{\sqrt{\frac{1 - \hat{r}_{XY}^2}{N - 2}}} \approx t_{N-2} \quad \hat{r}_{XY} - t_{\nu} \sigma_r \leq r_0 \leq \hat{r}_{XY} + t_{\nu} \sigma_r$$

Nota: Trattandosi di campioni normali, questo è anche un test di indipendenza stocastica di X da Y .

- Ipotesi $H_0: \rho = \rho_0$ (alle stesse condizioni): $\frac{\left(\frac{1}{2} \ln \frac{1 + \hat{r}_{XY}}{1 - \hat{r}_{XY}} - \frac{1}{2} \ln \frac{1 + \rho_0}{1 - \rho_0} \right)}{\frac{1}{\sqrt{N - 3}}} = \frac{\hat{Z} - \bar{Z}}{\sigma_Z} \approx z$

TEST NON – PARAMETRICI (distribution - free)

- **Test di buon adattamento (goodness of fit)**

Ipotesi H_0 : il campione è estratto da una popolazione con distribuzione teorica nota.

Dopo aver diviso un campione (con numerosità N) in m classi, si confrontano le frequenze assolute \hat{F}_i del campione con le probabilità p_i date dalla distribuzione teorica nota

$$\sum_{i=1}^m \frac{(\hat{F}_i - N p_i)^2}{N p_i} = N \sum_{i=1}^m \frac{(\hat{f}_i - p_i)^2}{p_i} \approx \chi_{m-h-1}^2$$

Nota: Il numero h rappresenta il numero di parametri di disturbo eventualmente determinati (solitamente 2). Il suddetto confronto si effettua, in modo analogo, anche per un campione bidimensionale (dove il numero dei parametri di disturbo è 5).

- **Test di indipendenza**

Ipotesi H_0 : i campioni X e Y sono indipendenti.

Date N coppie di valori (X, Y) , ordinate in una tabella a doppia entrata, suddivisa in m classi per quanto riguarda i valori X e n classi per quanto riguarda i valori Y , si confrontano le frequenze assolute, \hat{F}_{ij} con le frequenze assolute marginali $\hat{P}_i \hat{Q}_j$ (essendo i gdl: $\nu = nm - (n - 1) - (m - 1) - 1 = (n - 1)(m - 1)$):

$$\sum_{i=1}^m \sum_{j=1}^n \frac{\left(\hat{F}_{ij} - \frac{\hat{P}_i \hat{Q}_j}{N} \right)^2}{\frac{\hat{P}_i \hat{Q}_j}{N}} = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n \frac{(N\hat{F}_{ij} - \hat{P}_i \hat{Q}_j)^2}{\hat{P}_i \hat{Q}_j} = N \sum_{i=1}^m \sum_{j=1}^n \frac{(\hat{f}_i - \hat{p}_i \hat{q}_j)^2}{\hat{p}_i \hat{q}_j} \approx \chi_{(n-1)(m-1)}^2$$

□ Test di Kolmogorov-Smirnov

Ipotesi H_0 : il campione è estratto da una popolazione con distribuzione teorica nota.

Dopo aver diviso il campione (con numerosità N) in m classi, si confrontano le frequenze relative cumulate \hat{t}_i del campione con le probabilità cumulate P_i date dalla distribuzione teorica nota:

$$\max |\hat{t}_i - P_i| \approx D_N \quad ; \quad i = 1, N \quad \quad \text{Nota: } D_{.05} \xrightarrow{N > 100} \frac{1.36}{\sqrt{N}} \quad ; \quad D_{.01} \xrightarrow{N > 100} \frac{1.63}{\sqrt{N}}$$

Nota: A rigore il test dovrebbe essere eseguito considerando un dato alla volta; pertanto è opportuno che le eventuali classi siano le più piccole possibili. Inoltre il suddetto confronto si effettua anche per un campione bidimensionale e può essere usato anche il test d'indipendenza, effettuando i prodotti della frequenze marginali cumulate ed avendo cura di attenersi alle medesime precauzioni.

□ Test Pearson et al. per la normalità

Ipotesi H_0 : il campione è estratto da una popolazione con distribuzione normale.

Il test confronta i valori stimati dei coefficienti di asimmetria e curtosi con i valori teorici della distribuzione normale ($\bar{\gamma} = 0$; $\bar{\beta} = 3$):

$$\frac{\hat{\gamma}^2}{6/N} + \frac{(\hat{\beta} - 3)^2}{24/N} \approx \chi_2^2$$

□ Test di Mann-Whitney

Ipotesi H_0 : $\mu_X = \mu_Y$, per il confronto dei valori centrali di due campioni X e Y indipendenti.

I dati dei campioni vengono sostituiti dai corrispondenti ranghi i cui valori vanno da 1, per il dato di valore argomentale minimo, a $(N_X + N_Y)$, per il dato di valore argomentale massimo. Detta \hat{R}_X la somma dei ranghi del campione X , si ha:

$$\frac{\hat{R}_X - \frac{N_X(N_X + N_Y + 1)}{2}}{\sqrt{\frac{N_X N_Y (N_X + N_Y + 1)}{12}}} \approx z$$

□ Test di Siegel-Tukey

Ipotesi H_0 : $\sigma_X^2 = \sigma_Y^2$, per il confronto dei valori di dispersione di due campioni X e Y indipendenti.

I dati dei campioni vengono sostituiti dai corrispondenti ranghi i cui valori vanno da 1, per il dato il cui scarto in valore assoluto rispetto alla mediana è minimo, a $(N_X + N_Y)$, per il dato il cui scarto in valore assoluto rispetto alla mediana è massimo. Detta \hat{R}_X la somma dei ranghi del campione X , la conclusione del test è formalmente identica al precedente test di Mann-Whitney cui si rimanda, per quanto riguarda tanto il valore atteso, quanto la distribuzione normale di riferimento.

□ Test del segno (per i valori centrali)

Ipotesi H_0 : $\mu_X = \mu_Y$.

Nel caso dei cosiddetti studi "prima e dopo", cioè quando si misura due volte lo stesso campione, si

ottengono due campioni X (valore misurato "prima") e Y (valore misurato "dopo") non indipendenti. Per ogni coppia di valori argomentali, si determina il segno (+ o -), secondo la convenzione:

<i>valore "prima"</i>		<i>valore "dopo"</i>	<i>segno</i>
X	$>$	Y	-
X	$<$	Y	+
X	$=$	Y	<i>nessuno</i>

Detti: N_p = numero dei segni "più", N_m = numero dei segni "meno", $N_{tot} = N_p + N_m$ = numero totale dei

segni, e calcolato $\hat{f} = \frac{N_p}{N_{tot}}$ frazione dei segni "più" sul totale dei segni, si ha: $\frac{\hat{f} - 0.5}{\sqrt{\frac{0.5}{N_{tot}}}} \approx z$

□ **Test del segno (per i valori di dispersione)**

Ipotesi H_0 : $\sigma_X^2 = \sigma_Y^2$.

Anche in questo caso, si ottengono due campioni non indipendenti. Per ogni coppia di scarti in valore assoluto rispetto alla mediana, si determina il segno (+ o -), secondo la convenzione:

<i>valore "prima"</i>		<i>valore "dopo"</i>	<i>Segno</i>
$ X - med_X $	$>$	$ Y - med_Y $	-
$ X - med_X $	$<$	$ Y - med_Y $	+
$ X - med_X $	$=$	$ Y - med_Y $	<i>Nessuno</i>

Detti: N_p = numero dei segni "più" e $\hat{f} = \frac{N_p}{N_{tot}}$, si ha: $\frac{\hat{f} - 0.5}{\sqrt{\frac{0.5}{N_{tot}}}} \approx z$

Note: Per confrontare una frequenza relativa \hat{f} con una probabilità: $p = p_0$ (o due frequenze: \hat{f}_1, \hat{f}_2), si operano le seguenti standardizzazioni di cui quella usata per i test del segno costituisce un caso particolare:

$$\frac{\hat{f} - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}} \approx z \qquad \frac{\hat{f}_1 - \hat{f}_2}{\sqrt{\frac{\hat{f}_1(1-\hat{f}_1)}{N_1} + \frac{\hat{f}_2(1-\hat{f}_2)}{N_2}}} \approx z$$

La stessa standardizzazione è utilizzata per confrontare una contingenza \hat{c} data dalla differenza fra una frequenza relativa \hat{f} ed il prodotto delle corrispondenti frequenze relative marginali: $pq = p_0q_0$:

$$\frac{\hat{c}}{\sqrt{\frac{p_0q_0(1-p_0q_0)}{N}}} = \frac{\hat{f} - p_0q_0}{\sqrt{\frac{p_0q_0(1-p_0q_0)}{N}}} \approx z$$

□ **Test di Spearman**

Procedura per il calcolo del coefficiente di correlazione sui ranghi di Spearman:

1. ordinare i dati per ciascuna componente X e Y ;
2. assegnare i ranghi, separatamente, a ciascuna componente, nell'ordine crescente dei valori argomentali;
3. calcolare, elemento ad elemento, le differenze Δ_i fra i ranghi delle due componenti;

4. calcolare il coefficiente di correlazione sui ranghi di Spearman:
$$\hat{r}_{XY} = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^N \Delta_i^2$$

Ipotesi $H_0: \rho = 0$ (N deve essere sufficientemente grande):
$$\frac{\hat{r}_{XY}}{\sqrt{\frac{1 - \hat{r}_{XY}^2}{N - 2}}} \approx t_{N-2}$$

Ipotesi $H_0: \rho = \rho_0$:
$$\frac{\left(\frac{1}{2} \ln \frac{1 + \hat{r}_{XY}}{1 - \hat{r}_{XY}} - \frac{1}{2} \ln \frac{1 + \rho_0}{1 - \rho_0} \right)}{\frac{1}{\sqrt{N - 3}}} = \frac{\hat{Z} - \bar{Z}}{\sigma_Z} \approx z$$

Nota: Le v.c. χ^2 , t di Student e F di Fisher hanno le seguenti trasformazioni asintotiche alla v.c. normale:

$$\sqrt{2\chi_v^2} - \sqrt{v-1} \xrightarrow{v \rightarrow \infty} z \quad t_v \xrightarrow{v \rightarrow \infty} z \quad \frac{1}{2} \ln F_{v_1 v_2} \xrightarrow{v_1 v_2 \rightarrow \infty} z \quad \text{inoltre si ha: } F_{v_1 v_2}^{(-)} = \frac{1}{F_{v_2 v_1}^{(+)}}$$

□ Test di Welch e Tukey per il confronto di medie di campioni normali di diversa varianza

Ipotesi $H_0: \mu_X = \mu_Y$ ($\sigma_X^2, \sigma_Y^2 \neq \sigma_X^2$ incogniti): $\Delta \approx t_\nu$

Valore atteso: $\Delta = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\hat{\sigma}_X^2}{N_X} + \frac{\hat{\sigma}_Y^2}{N_Y}}}$ gradi di libertà: $\nu = \frac{\left(\frac{\hat{\sigma}_X^2}{N_X} + \frac{\hat{\sigma}_Y^2}{N_Y} \right)^2}{\left(\frac{\hat{\sigma}_X^2}{N_X} \right)^2 \frac{1}{N_X + 1} + \left(\frac{\hat{\sigma}_Y^2}{N_Y} \right)^2 \frac{1}{N_Y + 1}} - 2$

Nota: se il numero dei gradi di libertà ν non è un numero intero, esso viene arrotondato al numero intero più vicino, così da poter accedere alle usuali tabelle di distribuzione di probabilità.

BIBLIOGRAFIA DI RIFERIMENTO ²⁷

1. Baarda W. (1967): A Testing Procedure for Use in Geodetic Networks. Netherlands Geodetic Commission, *Publications on Geodesy New Series*, Delft, Vol. 2, n. 5.
2. Baarda W. (1967): Statistical Concepts in Geodesy. Netherlands Geodetic Commission, *Publications on Geodesy New Series*, Delft, Vol. 2, n. 4.
3. Barrodale I., Roberts F. (1973): An improved algorithm for discrete L_1 linear approximation. *SIAM Journal of Numerical Analysis*, n. 10 (1973).

²⁷ A conclusione del presente lavoro, è posto un doveroso commento, riguardo la sola presentazione di una bibliografia essenziale. Innanzitutto l'essenzialità della stessa risiede, oltreché nell'inevitabile autorevolezza degli autori citati, nell'essere stati gli stessi autori indispensabili e fondamentali per la formazione dell'autore del presente lavoro, spesso tramite anche una conoscenza personale ed, in ogni caso, con la lettura e lo studio di testi adatti agli scopi acclarati.

D'altra parte, la relativa semplicità del presente lavoro fa ritenere all'autore dello stesso, del tutto sufficiente la sola presentazione di una bibliografia essenziale. Infatti in questo caso specifico, il fornire un lungo elenco non avrebbe apportato informazioni veramente aggiuntive, rispetto a quanto qui riportato ed, in altra veste, esposto nei testi presentati come bibliografia essenziale (Consequentemente un lungo elenco sarebbe stato quantomeno prolisso e forse fuori luogo).

In ogni caso, per soddisfare ai comunque leciti desideri di un qualsiasi appassionato ed interessato lettore, si rimanda ai libri di *Trattamento delle Osservazioni* e, con una maggiore attenzione alle questioni di frontiera nella ricerca scientifica, agli atti delle *Associazioni Scientifiche Internazionali di settore*, ove rivolti al *Trattamento delle Osservazioni*. Da ultimo, anche la produzione scientifica dell'autore del presente lavoro può offrire spunti d'interesse.

Applicazioni d'interesse sono direttamente note e spaziano non solo, ovviamente, nel campo vastissimo delle discipline del rilevamento, ma anche in altri settori, molto diversi e differenti fra loro, quali, ad esempio, l'astronomia, la geologia tecnica, l'idrologia, l'ingegneria strutturale, i modelli urbanistici, la demografia, l'econometria e la biometria. In ogni caso, le metodologie e procedure presentate, unitamente agli algoritmi necessari per la loro applicazione, rispondono positivamente alla bisogna.

4. Barnett V., Lewis T. (1978): Outliers in statistical data. J. Wiley & Sons, New York.
5. Bartels R.H., Golub G.H. (1969): The simplex method of linear programming using LU-decomposition. In: *Communication of the ACM*, Vol. 12, n. 5, New York.
6. Bjerhammar A. (1973): Theory of Errors and Generalized Matrix Inverses. Elsevier Scientific Publishing Company, Amsterdam.
7. Bloomfield P., Steiger W.L. (1983): Least Absolute Deviations. Birkhäuser, Boston.
8. Cross P.A. (1982): Computer aided design of geodetic networks. In: *Int. Symp. on Geodetic Networks and Computations*, DGK Reihe B Heft 258/III, Monaco di Baviera.
9. Cross P.A. and Fagir A.H. (1982): Procedures for the first and second order design of vertical control networks, In: *Survey Control Networks*, HBSw, n. 7, Monaco di Baviera.
10. Dwyer P.S. (1951): Linear Computation. John Wiley & Sons, New York.
11. Forstner W. (1986): Reliability, Error detection and self calibration. In: *Int. Arch. of Photogram. Remote Sensing*, Rovaniemi, Vol. 26, Part 3/4.
12. Frederiksen P., Kubik K., Weng W. (1984): Oh, Grossororors!. In: *Int. Arch. of Photogram. Remote Sensing*, Rio de Janeiro, Vol. 25, Part A3a.
13. Fuchs H. (1982): Contributions to the adjustment by minimizing the sum of absolute residuals. In: *Manuscripta Geodaetica*, n. 7 (1982).
14. Grafarend E.W. et al. (1979): Optimierung Geodätischer Messoperationen. Wichmann, Kalsruhe.
15. Grafarend E.W., Sansò F. et al. (1985): Optimization and Design of Geodetic Networks. Springer, Berlino.
16. Hageman L.A., Young D.M. (1981): Applied Iterative Methods. Academic Press, New York.
17. Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. (1986): Robust Statistics. The Approach Based on Influence Functions. Wiley & Sons, New York.
18. Hawkins D.M. (1980): Identification of outliers. Chapman and Hall, Londra.
19. Huber P.J. (1964): Robust estimation of a location parameter. In: *Ann. Math. Statist.*, n. 35, p. 73–101.
20. Huber P.J. (1972): Robust statistics: a review. In: *Ann. Math. Statist.*, n. 43, p. 1041–1067.
21. Huber P.J. (1981): Robust Statistics. Wiley & Sons, New York.
22. Heiskanen W.A., Moritz H. (1967): Physical Geodesy. Freeman and Company, San Francisco.
23. Jorgensen P., Frederiksen P., Kubik K., Weng W (1984): Ah, robust estimation. In: *Int. Arch. of Photogram. Remote Sensing*, Rio de Janeiro, Vol. 25, part A3a.
24. Koch K.R. (1987): Parameter Estimation and Hypothesis Testing in Linear Models. Springer, Berlino.
25. Krarup T., Juhl J., Kubik K.: Götterdämmerung of least squares. In: *Int. Arch. of Photogram. Remote Sensing*, Hamburg, Vol. 23, Part B3.
26. Meissl P. (1982): Least Squares Adjustment a Modern Approach. Mitteilungen der geodätischen Institute der Technischen Universität Graz, Folge 43, Graz.
27. Mikhail E.M., Ackermann F. (1976): Observations and Least Squares. IEP–A Dun–Donnelley Publisher,, New York.
28. Moritz H. (1980): Advanced Physical Geodesy. Wichmann, Karlsruhe.
29. Muller H. (1984): A numerically efficient solution of the second order design problem. In: *Bulletin Géodésique*, n. 58 (1984).
30. Muller H. (1985): Second order design of combined linear–angular networks. In: *Bulletin Géodésique*, n. 59 (1985).
31. Rousseeuw P.J, Leroy A.M. (1987): Robust Regression and Outlier Detection. Wiley & Sons, New York.
32. Schaffrin B. (1981): Some proposals concerning the diagonal second order design of geodetic networks. In: *Manuscripta Geodaetica*, n. 6 (1981).
33. Schaffrin B. (1982): Some considerations on the optimal design of geodetic networks. In: *Int. Symp. on Geodetic Networks and Computations*, DGK Reihe B Heft 258/III, Monaco di Baviera.
34. Schaffrin B. (1985): On design problems in geodesy using models with prior information. In: *Statistics & Decision*, Suppl. Issue n. 2 (1985).
35. Schmitt G. (1977): Experiences with the second order design of geodetic networks. In: *Proc. Int. Symp. on Optimization of Design and Computation of Control Networks*, Sopron.
36. Schmitt G. (1978): Numerical problems concerning the second order design of geodetic networks. In: *Proc. 2nd Int. Symp. Related to the Redefinition of North American Geodetic Networks*, Washington DC:
37. Schmitt G. (1987): Optimal design problems. In: *Geodetic Theory and Methodology*, University of Calgary, Calgary.
38. Stefanovic P. (1978): Blunders and least squares. ITC Journal, n. 1 (1978).
39. Thompson E.H. (1969): An Introduction to be Algebra of Matrices With Some Applications. A. Hilger, London.
40. Wimmer H. (1982): Second order design of geodetic networks by an interactive approximation of a given criterion matrix. In: *Int. Symp. on Geodetic Networks and Computations*, DGK Reihe B Heft 258/III, Monaco di Baviera.