

UNIVERSITÀ CATTOLICA DEL SACRO CUORE DI MILANO

Interfacoltà Economia/Scienze Bancarie, Finanziarie e Assicurative

Corso di Laurea Specialistica in Scienze Statistiche ed Economiche

Indirizzo per il Controllo della Qualità ed il Marketing

**Modelli di regressione per fattori controllabili
non lineari nei parametri:
possibili contributi della Geometria Differenziale**

**Tesi di Laurea di
Federico ANDREIS
Matr. N°. 3307235**

Relatore: Chiar.mo Prof. Angelo Zanella

Anno Accademico 2005/2006

«Più si sviluppa una teoria matematica, più la sua costruzione
procede armoniosamente e uniformemente, e relazioni insospettate
si scoprono tra branche della scienza fino ad allora separate.»

*David Hilbert (1862-1943), in N. Rose, **Mathematical Maxims and Minims.***

Indice

Capitolo I

Introduzione.....	6
1.1 Regressione lineare e metodo dei minimi quadrati.....	8
1.2 Intepretazione geometrica del modello di regressione lineare.....	11
1.3 Il modello di regressione lineare nei parametri.....	23
1.4 Stima dei parametri di un modello lineare.....	27
1.5 Approssimazione mediante linearizzazione.....	36
1.6 Intenti del lavoro.....	39

Capitolo II

Breve introduzione al calcolo differenziale assoluto (calcolo tensoriale).....	40
2.1 Rappresentazione cartesiana dei vettori.....	40
2.2 Invarianza delle relazioni vettoriali.....	42
2.3 Rappresentazione cartesiana dei tensori doppi.....	42
2.4 Rappresentazione cartesiana dei tensori multipli.....	44
2.5 Richiami di Geometria Differenziale: spazi euclidei e varietà curve.....	49
Superficie.....	53
2.6 Coordinate curvilinee o generali.....	55
Coordinate polari.....	57
Coordinate cilindriche.....	59
2.7 Metrica di uno spazio curvo e tensore metrico fondamentale.....	61
2.8 Interpretazione di un modello non lineare come una varietà riemanniana.....	70
Modello lineare.....	71
Modello non lineare.....	73

Capitolo III

Il concetto di curvatura.....	75
Curvatura intrinseca ed estrinseca.....	75
3.1 Curvatura di una linea nel piano.....	76
Il cerchio osculatore.....	77
3.2 Curvatura di una linea nello spazio.....	79

Prima curvatura generalizzata.....	81
Seconda curvatura generalizzata.....	82
3.3 Curvatura di una superficie, Prima Forma Fondamentale.....	84
Curvatura della superficie cilindrica.....	84
Prima Forma Fondamentale.....	87
3.4 Operatore di Weingarten e Seconda Forma Fondamentale.....	90
 Capitolo IV	
La curvatura nei modelli di regressione non lineari nei parametri.....	97
4.1 Scomposizione del vettore curvatura.....	99
4.2 Perché introdurre misure di curvatura nell'analisi dei modelli non lineari.....	100
Non linearità intrinseca.....	103
Non linearità indotta dai parametri.....	103
La misura di distorsione di Box.....	104
Lo Standard Radius.....	105
4.3 Costruzione delle misure di non linearità.....	106
Le tecniche di approssimazione nella stima.....	106
Costruzione degli indici.....	108
Effetti della replicazione di un disegno sperimentale sulla curvatura.....	118
Diverse implicazioni della curvatura intrinseca e di quella indotta dai parametri.....	119
4.4 Misure relative di non linearità.....	120
Calcolo delle misure di curvatura relative.....	122
Interpretazione geometrica delle componenti dell'array \ddot{A}^T	124
Calcolo della misura di distorsione di Box.....	127
Ulteriori considerazioni sulle implicazioni della non linearità intrinseca.....	129
Il t -value e le misure di curvatura di Bates e Watts.....	131
Considerazioni sulla correlazione fra le stime LS in un modello non lineare.....	132
Benefici della riparametrizzazione.....	132
4.5 Studio tramite simulazioni del comportamento delle stime secondo i LS.....	134
Un semplice modello non lineare monoparametrico.....	135
Il modello a due parametri di Michaelis-Menten.....	147
Un modello di tipo Weibull a quattro parametri.....	159

Conclusioni	169
Ringraziamenti	171
Appendice	175
A.1 La fattorizzazione QR	175
A.2 Coincidenza della misura di Bates e Watts con la curvatura normale	177
A.3 Listati del codice impiegato per le simulazioni	183
Bibliografia	196

CAPITOLO I

Introduzione

Nell'ambito delle scienze sperimentali un aspetto fondamentale della ricerca coinvolge l'individuazione di relazioni all'interno di un insieme di variabili delle quali almeno una sia da ritenersi casuale, sia essa per fluttuazioni aleatorie, possibili errori di misurazione o effettiva impossibilità di effettuare detta rilevazione.

La presenza di una componente stocastica richiede necessariamente l'abbandono dell'ottica deterministica (o "Galileiana"), la quale condurrebbe alla specificazione di soli legami funzionali esatti tra le variabili atte a descrivere la realtà fenomenica oggetto di studio, e della quale ci si potrà avvalere soltanto in un primo momento con lo scopo di poter meglio orientare il lavoro, senza avere però alcuna pretesa di completezza di analisi in un ambito di incertezza.

Tenuto conto dell'impossibilità sperimentale di stabilire e controllare esattamente tutte le variabili, l'approccio deterministico necessiterà dunque di essere completato e corretto. Tale compito viene svolto avvalendosi di quelle tecniche statistiche che consentono di esprimere le relazioni intercorrenti nell'insieme di variabili tramite espressioni funzionali delle stesse, che si riterranno generalmente valide a meno di una componente di errore, tenendo conto della quale saremo in grado di non tralasciare l'aleatorietà onnipresente nelle situazioni reali.

Nei problemi di regressione tipicamente viene individuata una delle variabili, sia essa y , spesso chiamata **risposta** o **variabile dipendente**, di particolare interesse. Le altre variabili, siano esse x_1, x_2, \dots, x_n , solitamente indicate come **esplicative**, **regressori** o **variabili indipendenti**, sono principalmente impiegate per spiegare o predire il comportamento di y .

Nel caso in cui l'osservazione del fenomeno o, ad esempio, la rappresentazione grafica dei dati a nostra disposizione suggerisca una qualche relazione (un

andamento più o meno regolare nella nube dei punti) tra la variabile dipendente e una o più di quelle indipendenti, allora è nostra speranza quella di poter esprimere questo legame tramite una qualche funzione matematica f di modo che:

$$y \approx f(x_1, x_2, \dots, x_n, \vartheta_0, \vartheta_1, \dots, \vartheta_m)$$

rappresenti una approssimazione accettabile di detto legame, dove $\vartheta_0, \vartheta_1, \dots, \vartheta_m$ saranno incogniti (ma costanti) coefficienti, o parametri, che dovranno essere stimati sulla base dei dati in nostro possesso tramite una qualche procedura statistica, mentre x_1, x_2, \dots, x_n sono i regressori.

Supponendo di avere ottenuto una f che sia capace di descrivere il fenomeno con un livello di accuratezza sufficientemente elevato, potremmo dunque predire y per un dato insieme di x . Ad esempio, y potrebbe rappresentare il prezzo di una autovettura usata di una determinata marca, x_1 il numero di precedenti proprietari, x_2 l'età della vettura e x_3 il chilometraggio. Come già introdotto in precedenza, una simile relazione funzionale non potrà mai valere con esattezza, poiché i dati conterranno immancabilmente fluttuazioni inspiegabili, così come sarà solitamente presente un certo grado di errore di misurazione (come ad esempio un conteggio non accurato dei chilometri percorsi).

Le variabili esplicative potranno essere casuali a loro volta oppure fissate (controllate). Si consideri un esperimento condotto per misurare la resa (y) di grano rispetto a molteplici differenti livelli di densità della semina (x_1) ed impiego di fertilizzanti (x_2). In questo caso ambedue i regressori sono fissati (si parlerà di **fattori controllabili**). Se, al momento della semina, il pH del suolo (x_3) venisse anch'esso misurato, esso costituirebbe un regressore casuale. Nel presente lavoro l'interesse sarà rivolto a quei soli modelli di regressione che considerino fattori controllabili, ovvero non stocastici. In particolare ci si riferirà all'insieme dei valori fissati per i fattori controllabili come al **disegno sperimentale** dello studio.

1.1 Regressione lineare e metodo dei minimi quadrati

Il termine “modello di regressione lineare” viene talvolta impiegato per indicare due oggetti differenti: il primo impiego si riferisce ad una relazione del tipo “linea retta” (od eventuali estensioni multidimensionali, parliamo di varietà lineari) fra due (o più) variabili, mentre il secondo si riferisce a modelli in cui i **parametri**, ovvero le quantità che devono essere stimate, compaiono nell’espressione linearmente; quest’ultima sarà l’accezione di linearità adottata in questo lavoro.

Ad esempio, i seguenti sono modelli di regressione lineari nei parametri:

$$y = \mathcal{G}_0 + \mathcal{G}_1 x_1 + \varepsilon$$

$$y = \mathcal{G}_0 + \mathcal{G}_1 x_1 + \mathcal{G}_2 x_1^2 + \varepsilon$$

$$y = \mathcal{G}_0 + \mathcal{G}_1 x_1 + \mathcal{G}_2 x_2 + \varepsilon$$

dove le lettere greche $\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_m$ e le x_1, x_2, \dots, x_n assumono i significati già esposti. Il primo modello è un modello “linea retta”, il secondo è una parabola, il terzo è il più semplice esempio di regressione lineare multipla, e geometricamente rappresenta l’equazione di un piano; ε indica quella componente di errore, sempre presente nei fenomeni oggetto di studio, che sfugge ad una descrizione deterministica, sotto le opportune condizioni che sia rappresentativa di una variabile casuale normale, di media nulla e varianza finita e costante (condizioni queste che si riterranno sempre verificate nell’ambito di questo lavoro per comodità e per le pregevoli proprietà relative agli stimatori cui esse conducono).

La stima degli incogniti parametri dei modelli sopraccitati viene effettuata tramite la tecnica dei **Minimi Quadrati** (Least Squares, LS d’ora in avanti), la quale richiede la soluzione di un problema di minimo nelle incognite $\mathcal{G}_0, \mathcal{G}_1, \dots, \mathcal{G}_m$ di una opportuna funzione di scarto quadratica; gli stimatori ottenuti attraverso l’applicazione di questo metodo, sotto le ipotesi sulla componente di errore di normalità, media nulla e varianza finita e costante, coincidono con quelli di **Massima Verosimiglianza** (Maximum Likelihood, ML d’ora in avanti), che sono

noti essere non distorti e di minima varianza nella classe degli stimatori lineari non distorti. In particolare lo stimatore risulterà avere anche distribuzione normale, in quanto combinazione lineare di variabili casuali gaussiane (in virtù della normalità dell'errore, la variabile dipendente avrà anch'essa distribuzione normale), informazione fondamentale ai fini della necessaria inferenza statistica. Dunque se soddisfatte le ipotesi avanzate, il criterio dei LS fornisce i migliori stimatori possibili disponibili nella pratica.

Presentando formalmente il principio dei minimi quadrati possiamo quindi dire che: la metodologia di stima LS è impiegata nell'analisi di un insieme di dati che consiste di risposte $y_t, t=1, \dots, n$, ottenute in corrispondenza di prefissate condizioni sperimentali (fattori) $\mathbf{x}_t, t=1, \dots, n$. Si suppone che la relazione intercorrente tra le risposte e i regressori possa essere adeguatamente rappresentata da un'equazione del tipo

$$y_t = f(\mathbf{x}_t, \boldsymbol{\theta}) + \varepsilon_t$$

in cui $\boldsymbol{\theta} = (\vartheta_1, \vartheta_2, \dots, \vartheta_p)'$ è l'insieme degli ignoti parametri oggetto di stima, mentre $\varepsilon_t, t=1, \dots, n$, è la componente erratica (supposta additiva) casuale dovuta ad errori di misurazione. Se gli errori soddisfano le condizioni

$$\begin{aligned} E[\varepsilon_t] &= 0, & t &= 1, \dots, n \\ E[\varepsilon_t, \varepsilon_s] &= \sigma^2 \delta_{ts}, & t, s &= 1, \dots, n \end{aligned}$$

dove δ_{ts} è il delta di Kronecker (che assume valore 1 quando $t = s$, 0 altrimenti), allora il valore $\hat{\boldsymbol{\theta}}$ che minimizza la somma dei quadrati dei residui

$$S(\boldsymbol{\theta}) = \sum_{t=1}^n (y_t - f(\mathbf{x}_t, \boldsymbol{\theta}))^2$$

viene chiamata **stima di θ secondo il criterio dei LS**. Se gli errori sono normalmente distribuiti, allora $\hat{\theta}$ coincide con la stima di ML di θ .

Il vettore parametrico ed il disegno sperimentale determinano il valore della funzione del modello $f(\mathbf{x}, \theta)$ il quale, sotto l'assunzione di nullità in media degli errori, coincide con il valore atteso di y condizionatamente ad \mathbf{x} e θ

$$E[y | \mathbf{x}, \theta] = f(\mathbf{x}, \theta).$$

Per ogni condizione sperimentale \mathbf{x}_t , consideriamo la **risposta attesa condizionata**

$$\eta_t(\theta) = E[y_t | \theta] = f(\mathbf{x}_t, \theta),$$

possiamo dunque riscrivere la somma dei quadrati come segue:

$$S(\theta) = \|\mathbf{y} - \boldsymbol{\eta}(\theta)\|^2$$

dove $\boldsymbol{\eta}(\theta) = (\eta_1(\theta), \eta_2(\theta), \dots, \eta_n(\theta))'$, e con $\|\cdot\|$ indichiamo la norma (lunghezza) di un vettore. In questa forma $S(\theta)$ ha l'interpretazione geometrica di quadrato della distanza tra i vettori \mathbf{y} ed $\boldsymbol{\eta}(\theta)$ in uno spazio campionario n -dimensionale.

Al variare di θ il vettore $\boldsymbol{\eta}(\theta)$ traccia una superficie p -dimensionale, chiamata **luogo delle soluzioni**, immersa nello spazio campionario n -dimensionale. In generale \mathbf{y} non sarà un punto appartenente a tale insieme, e la stima secondo LS $\hat{\theta}$ è quel vettore parametrico tale per cui $\boldsymbol{\eta}(\hat{\theta})$ è il punto sul luogo delle soluzioni più vicino ad \mathbf{y} . Si fa notare che il luogo delle soluzioni non dipende da \mathbf{y} , e quindi sarà possibile esaminarlo per ipotetici esperimenti senza avere la necessità di simulare le risposte.

Nei paragrafi seguenti introdurremo l'interpretazione geometrica del modello di regressione lineare dei parametri, punto concettuale di fondamentale importanza per

la seguente estensione al caso non lineare, definendo rigorosamente dal punto di vista geometrico il problema di stima parametrica secondo i LS e la sua soluzione.

1.2 Interpretazione geometrica del modello di regressione lineare

Quanto segue è tratto dalle dispense del corso di Programmazione degli Esperimenti del professor Angelo Zanella (pagine 29-49).

Uno spazio Euclideo \mathbb{R}^n a n dimensioni si definisce come un insieme di tutte le possibili n -uple ordinate di numeri reali, quindi un punto di tale spazio è una n -upla di valori (x_1, x_2, \dots, x_n) , rappresentabile in notazione matriciale mediante un vettore colonna $\mathbf{x} = (x_1, x_2, \dots, x_n)'$.

Si osservi che per definire uno spazio metrico bisogna introdurre la nozione di distanza tra due punti. Ad esempio gli spazi probabilistici sono, in sostanza, spazi non metrici, poiché in essi non si definisce, in generale, la distanza tra due punti.

La distanza nello spazio euclideo è l'estensione di quella euclidea per lo spazio ordinario.

Se si considerano cioè due n -uple $x_{1i}, x_{2i}, i = 1, \dots, n$, ovvero due vettori $\mathbf{x}_1, \mathbf{x}_2$, per distanza tra due punti si intende la seguente espressione:

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} = \|\mathbf{x}_1 - \mathbf{x}_2\|$$

L'espressione $\|\mathbf{x}_1 - \mathbf{x}_2\|$ indica la **norma** della differenza dei due vettori rappresentanti punti dello spazio ad n dimensioni.

La definizione di tale distanza porta ad alcune proprietà interessanti.

1. E' definibile un ente che è l'esatto corrispondente n -dimensionale di una retta. Si può infatti definire un sottoinsieme dello spazio \mathbb{R}^n con coordinate parametriche dei suoi elementi date da:

$$\Gamma : \begin{cases} \eta_1 = x_{10}t \\ \eta_2 = x_{20}t \\ \dots \\ \eta_n = x_{n0}t \end{cases} \quad \text{o equivalentemente} \quad \Gamma : \boldsymbol{\eta} = \mathbf{x}_0 t$$

Dove t è una variabile reale e $x_{10}, x_{20}, \dots, x_{n0}$ sono delle costanti assegnate. Se ci si limita a considerare le prime due equazioni di questa rappresentazione si vede che esse rappresentano, in \mathbb{R}^n , l'equazione parametrica di una retta passante per l'origine. L'ente appena definito possiede una proprietà fondamentale che lo fa corrispondere, in \mathbb{R}^n , alla retta intuitiva. Precisamente, facendo riferimento allo spazio ordinario, se si considerano due punti P_1 e P_2 appartenenti ad una retta per l'origine, la lunghezza dell'"arco" di retta da loro definito è pari alla differenza delle distanze di P_1 e P_2 dall'origine.

Si vede subito essere questa una proprietà intrinseca propria della retta poiché, se si considera un'altra linea per l'origine, non si riesce, in base alla definizione di distanza, ad esprimere la lunghezza dell'arco fra due punti in questo modo.

Considerando ora un ente del tipo Γ in \mathbb{R}^n , la lunghezza l di un arco \mathcal{Y} fra due punti, definita come:

$$l = \int_{\mathcal{Y}} ds(x)$$

dove ds rappresenta la distanza tra due punti infinitamente prossimi, si può vedere come in base alla definizione di distanza data in precedenza può essere ancora espressa come differenza delle distanze, rispetto all'origine, dei due punti considerati.

In altri termini, se a partire da P_1 si va a P_2 lungo la linea Γ , si trova che il valore del precedente integrale è proprio uguale alla distanza tra i due punti. Tale ente estende dunque, ad \mathbb{R}^n , la nozione di retta.

2. Si può anche estendere la definizione di angolo tra due vettori definendo il suo coseno per mezzo della disuguaglianza di Cauchy-Schwartz.

Se $\mathbf{\eta}_1 = (\eta_{11}, \eta_{12}, \dots, \eta_{1n})'$, $\mathbf{\eta}_2 = (\eta_{21}, \eta_{22}, \dots, \eta_{2n})'$, sono due vettori di \mathbb{R}^n , detta disuguaglianza stabilisce che:

$$\|\mathbf{\eta}_1' \mathbf{\eta}_2\| = \left(\sum_{j=1}^n \eta_{1j} \eta_{2j} \right)^{1/2} \leq \|\mathbf{\eta}_1\| \cdot \|\mathbf{\eta}_2\|.$$

Può pertanto definirsi una quantità $-1 \leq \cos \theta \leq 1$ come segue:

$$\mathbf{\eta}_1' \mathbf{\eta}_2 = \|\mathbf{\eta}_1\| \cdot \|\mathbf{\eta}_2\| \cos \theta.$$

In corrispondenza si definirà quale valore dell'angolo formato dai due vettori (i quali, ricordiamo, possono sempre pensarsi applicati in uno stesso punto preso come origine del sistema di riferimento), quello della funzione

$$\theta = \cos^{-1} \left[\frac{\sum_{j=1}^n \eta_{1j} \eta_{2j}}{\left(\sum_{j=1}^n \eta_{1j}^2 \right)^{1/2} \left(\sum_{j=1}^n \eta_{2j}^2 \right)^{1/2}} \right],$$

con $0 \leq \theta \leq \pi$, dove \cos^{-1} indica la funzione inversa del coseno. I due vettori avranno la stessa direzione e verso se $\theta = 0$, stessa direzione ma verso opposto se $\theta = \pi$. Vettori aventi la stessa direzione verranno detti **paralleli**. L'espressione precedente consente anche di estendere la nozione di ortogonalità. Due vettori non nulli di \mathbb{R}^n verranno detti ortogonali se per il loro prodotto scalare sussiste la condizione:

$$\mathbf{\eta}_1' \mathbf{\eta}_2 = \sum_{j=1}^n \eta_{1j} \eta_{2j} = 0$$

il che implica che sia $\theta = \pi/2$.

Se anche in \mathbb{R}^n si interpreta un vettore come un “segmento di retta orientato”, è immediato definire l’angolo formato da due rette, in prima istanza passanti per l’origine degli assi di riferimento.

Per quanto già detto le equazioni delle due rette espresse in forma parametrica definiscono due vettori nell’origine: $\boldsymbol{\eta}_1 = (x_{11}t, x_{12}t, \dots, x_{1n}t)'$, $\boldsymbol{\eta}_2 = (x_{21}t, x_{22}t, \dots, x_{2n}t)'$.

In virtù dell’espressione

$$\theta = \cos^{-1} \left[\frac{\sum_{j=1}^n \eta_{1j} \eta_{2j}}{\left(\sum_{j=1}^n \eta_{1j}^2 \right)^{1/2} \left(\sum_{j=1}^n \eta_{2j}^2 \right)^{1/2}} \right]$$

siamo in grado di definire l’angolo formato da tali vettori, il quale, non dipendendo da t , verrà assunto come angolo formato dalle due rette. Ne risulta la definizione di rette **parallele** e di rette **ortogonali** esattamente come si è fatto per i vettori. Tali nozioni sono poi estendibili in modo immediato anche al caso di rette qualsiasi, non passanti per l’origine, a patto di considerare una loro traslazione che ne induca, appunto, il passaggio per l’origine.

La nozione di parallelismo consente, anche in \mathbb{R}^n , di interpretare il vettore somma di due vettori come ottenuto mediante la **regola del parallelogramma**. Quella di ortogonalità consente, inoltre, di introdurre in n dimensioni un sistema di assi cartesiani ortogonali, cioè costituito da n rette orientate, mutualmente ortogonali. Siano $\boldsymbol{a}_i, i=1,2,\dots,n$ i corrispondenti versori, ovvero vettori di modulo unitario diretti ed orientati come gli assi e, quindi, mutualmente ortogonali. Si ricordi che dati n vettori $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ di \mathbb{R}^n , linearmente indipendenti, si definiscono “coordinate di un punto”, cui corrisponde il vettore $\boldsymbol{v} = (v_1, v_2, \dots, v_n)'$, rispetto alla **base** costituita da $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$, la n -upla di valori $\beta_1, \beta_2, \dots, \beta_n$ tali che:

$$\boldsymbol{v} = \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \dots + \beta_n \boldsymbol{x}_n$$

certamente esistente per ipotesi. Gli assi coordinati sono rette orientate caratterizzate dal fatto che i loro punti hanno tutte le coordinate nulle meno una e, quindi, coincidono con le rette definite dai vettori $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. In particolare, quindi, gli assi coordinati che assegnano ad un punto le coordinate (v_1, v_2, \dots, v_n) devono avere versori $\mathbf{a}_i = (0, \dots, 1, \dots, 0)'$, devono cioè presentare solo la i -esima componente diversa da 0 e di valore unitario.

Detti versori sono, poi, ovviamente linearmente indipendenti e mutualmente ortogonali. Segue quindi anche per un vettore $\mathbf{v} = (v_1, v_2, \dots, v_n)'$ di \mathbb{R}^n la rappresentazione come combinazione lineare dei versori:

$$\mathbf{v} = v_1 \mathbf{a}_1 + v_2 \mathbf{a}_2 + \dots + v_n \mathbf{a}_n$$

Con coefficienti uguali alle corrispondenti componenti di \mathbf{v} , che possono così interpretarsi quali “coordinate cartesiane ortogonali” del punto di \mathbb{R}^n immagine del “vertice” del vettore. Una base di vettori ortogonali di modulo unitario viene denominata **base ortonormale**.

3. Nello spazio euclideo ordinario sussiste l'estensione bidimensionale della retta, ovvero il piano. In \mathbb{R}^n , $n > 3$, si può introdurre, come visto sopra, la nozione di retta, come insieme unidimensionale, che conserva formalmente le proprietà essenziali della retta; il procedimento può estendersi introducendo insiemi bidimensionali e tridimensionali che conservano le proprietà formali dei piani e dell'intero spazio tridimensionale. Per $n > 4$ il procedimento può estendersi ulteriormente con la definizione di sottoinsiemi a $4, \dots, n-1$ dimensioni che hanno le stesse proprietà di $\mathbb{R}^4, \dots, \mathbb{R}^{n-1}$. Tali estensioni possono tutte ricondursi ad un'unica nozione, quella di **varietà lineare omogenea** di \mathbb{R}^n . Per varietà lineare omogenea V di \mathbb{R}^n si intende un insieme $V \subset \mathbb{R}^n$ i cui elementi o punti sono costituiti da n -uple di valori $\eta_1, \eta_2, \dots, \eta_n$, cui si riterrà associato il vettore $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)'$, esprimibile nel modo seguente:

$$\begin{cases} \eta_1 = x_{11}\beta_1 + x_{21}\beta_2 + \dots + x_{m1}\beta_m \\ \eta_2 = x_{12}\beta_1 + x_{22}\beta_2 + \dots + x_{m2}\beta_m \\ \dots \\ \eta_n = x_{1n}\beta_1 + x_{2n}\beta_2 + \dots + x_{mn}\beta_m \end{cases}$$

dove $x_{ji}, j = 1, 2, \dots, m; i = 1, 2, \dots, n$ sono delle costanti reali assegnate, $\beta_j, j = 1, 2, \dots, m$ sono variabili reali che possono assumere ciascuna valori in $(-\infty, \infty)$. Si noti come se nella rappresentazione precedente si considera solo il primo addendo di ogni equazione ci si riconduce alle equazioni definenti una retta per l'origine in \mathbb{R}^n ; se si considerano solo i primi due addendi si dimostra che si ottiene un ente che rappresenta, in generale, l'estensione del piano intuitivo in \mathbb{R}^n , e così via.

Si può dare, grazie alla notazione matriciale, una rappresentazione più compatta del sistema di equazioni di cui poco sopra, come relazioni fra vettori colonna, ovvero:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

dove $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)'$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_m)'$ e \mathbf{X} è una matrice di dimensione $n \times m$ tale che:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{m1} \\ x_{12} & x_{22} & \dots & x_{m2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \dots & x_{mn} \end{bmatrix}.$$

La dimensione della varietà è data dal rango o caratteristica k della matrice \mathbf{X} , che indica il numero di colonne della stessa linearmente indipendenti. In corrispondenza indicheremo con il simbolo $V_{(k)}$ la varietà lineare di dimensione, appunto, k .

(a) Ovviamente $k \leq n$, quindi se $n < m$ sussistono almeno $(m-n)$ colonne che possono esprimersi come combinazione lineare delle restanti n ; ridefinendo i parametri β ci si potrà, quindi, sempre ricondurre al caso $m \leq n$.

(b) Se, essendo $m = n$, si ha pure $k = m = n$, vale a dire che la matrice, ora ritenuta quadrata, ha rango completo, allora il sistema di equazioni introdotto definisce, al variare dei parametri $\beta_j, j = 1, 2, \dots, n$, l'intero spazio \mathbb{R}^n e non si ha una varietà lineare in senso proprio. In tale caso, infatti, la matrice \mathbf{X} ammette un'inversa \mathbf{X}^{-1} e $\forall \boldsymbol{\eta} \in \mathbb{R}^n$ si possono trovare dei valori $\beta_j^*, j = 1, 2, \dots, n$, che consentono di esprimere $\boldsymbol{\eta}$ nella forma del sistema di equazioni sopra presentato. A tal fine basta infatti prendere

$$\boldsymbol{\beta}^* = \mathbf{X}^{-1} \boldsymbol{\eta}$$

$\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*, \dots, \beta_n^*)'$. Se introduciamo i vettori $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ con elementi costituiti dalle corrispondenti colonne di \mathbf{X} , si può anche scrivere:

$$\boldsymbol{\eta} = \beta_1^* \mathbf{x}_1 + \beta_2^* \mathbf{x}_2 + \dots + \beta_n^* \mathbf{x}_n.$$

Nel caso in discussione avremo dunque che i vettori $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ costituiscono una **base** per \mathbb{R}^n , cui è associato il sistema di coordinate $\beta_1^*, \beta_2^*, \dots, \beta_n^*$. Se i vettori $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ non sono tutti mutuamente ortogonali allora al riguardo si parla di “sistema di coordinate oblique”.

(c) Se $k \leq m < n$ allora le colonne della matrice definiscono un sottoinsieme proprio di \mathbb{R}^n di dimensione k . Se $k=m$ si dirà che la varietà ha **rango pieno o completo**. Concettualmente, come nel caso (a), se $k < m$ si potranno esprimere $(m-k)$ colonne di \mathbf{X} come combinazioni lineari delle restanti k e ricondursi al caso di rango pieno

attraverso una ridefinizione dei parametri β . Con riferimento a questo caso, in analogia al precedente, si può scrivere:

$$\boldsymbol{\eta} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_k \mathbf{x}_k$$

i vettori $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ costituiscono una **base** per $V_{(k)}$, cui risulta associato il sistema di coordinate, in generale oblique, $\beta_1, \beta_2, \dots, \beta_k$.

Si consideri una varietà lineare $V_{(k)}$ di \mathbb{R}^n , $k < n$, e sia $\mathbf{y} \in \mathbb{R}^n, \mathbf{y} \notin V_{(k)}$ un vettore di \mathbb{R}^n non appartenente alla varietà. Si dirà che \mathbf{y} è **ortogonale** alla varietà $V_{(k)}$, indicando tale proprietà con il simbolo $\mathbf{y} \perp V_{(k)}$, se:

$$\mathbf{y}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} = \mathbf{0}, \quad \forall \boldsymbol{\beta}.$$

Vale a dire se \mathbf{y} è ortogonale ad ogni vettore definito da un punto (di coordinate $\mathbf{X}\boldsymbol{\beta}$) della varietà.

Abbiamo in corrispondenza di queste considerazioni il seguente teorema:

Teorema 1.2.1 (della proiezione)

Sia $V_{(k)}$ una varietà lineare di \mathbb{R}^n con $1 \leq k < n$ e $\mathbf{y} \in \mathbb{R}^n, \mathbf{y} \notin V_{(k)}$ un vettore di \mathbb{R}^n non appartenente alla varietà.

(a) Si può sempre scomporre \mathbf{y} nella somma di due vettori \mathbf{w} e \mathbf{z} tali che:

$$\mathbf{y} = \mathbf{w} + \mathbf{z}$$

e $\mathbf{w} \in V_{(k)}, \mathbf{z} \notin V_{(k)} \wedge (\mathbf{z} = \mathbf{y} - \mathbf{w}) \perp V_{(k)}$, cioè \mathbf{w} appartenga alla varietà $V_{(k)}$ mentre \mathbf{w} non le appartiene ed è ortogonale a $V_{(k)}$.

(b) Tale scomposizione è unica.

(c) Il vettore \mathbf{w} è tale che:

$$\begin{aligned} \mathbf{z}'\mathbf{z} &= (\mathbf{y} - \mathbf{w})'(\mathbf{y} - \mathbf{w}) = \|\mathbf{y} - \mathbf{w}\|^2 = \\ &= \inf_{\mathbf{u} \in V_{(k)}} [(\mathbf{y} - \mathbf{u})'(\mathbf{y} - \mathbf{u})]. \end{aligned}$$

Questo risultato estende in \mathbb{R}^n l'esistenza e l'unicità della proiezione ortogonale di un punto di \mathbb{R}^3 su una retta o un piano cui non appartiene. Il vettore \mathbf{w} con le proprietà stabilite dal precedente teorema viene in corrispondenza indicato come **proiezione** ortogonale di \mathbf{y} su $V_{(k)}$.

Osservazioni

(a) Si dimostra agevolmente che nel caso $k=1$, che corrisponde alla proiezione di un vettore su una retta, risulta:

$$\mathbf{w} = \frac{\mathbf{y}'\mathbf{x}_1}{\|\mathbf{x}_1\|^2} \mathbf{x}_1 = \mathbf{y}' \left(\frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} \right) \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} = (\mathbf{y}'\boldsymbol{\alpha})\boldsymbol{\alpha}$$

dove $\|\cdot\|$ indica il modulo di un vettore, e quindi $\boldsymbol{\alpha}$ è il versore di \mathbf{x}_1 . Nel caso in oggetto la proiezione di \mathbf{y} coincide con il vettore diretto come $\boldsymbol{\alpha}$, con modulo pari al valore assoluto del prodotto scalare $\mathbf{y}'\boldsymbol{\alpha}$ e verso stabilito dal segno di detto prodotto. In base alla definizione del coseno dell'angolo θ formato da due vettori di \mathbb{R}^n introdotta in precedenza possiamo scrivere equivalentemente:

$$\mathbf{w} = [\|\mathbf{y}\| \cos\theta] \mathbf{a}.$$

(b) In \mathbb{R}^n oltre alla base ortonormale di cui si è parlato prima (che definiamo **fondamentale**, poiché composta dai versori fondamentali di \mathbb{R}^n), se ne possono definire infinite altre; siano $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \dots, \tilde{\mathbf{a}}_n$ i versori che definiscono una generica di tali basi. In corrispondenza resta definito un nuovo sistema di coordinate per un generico punto P di \mathbb{R}^n e, quindi, di valori per le componenti del vettore \mathbf{V} ad esso associato (che dal punto di vista geometrico, come segmento orientato di determinata lunghezza, rimane invariato quale che sia il punto di applicazione) avente come componenti le coordinate del punto P . Anche le nuove coordinate saranno dette “cartesiane ortogonali”. Il legame fra le nuove componenti (coordinate) $\tilde{v}_j, j=1,2,\dots,n$, e le originarie, rispetto al sistema fondamentale, $v_j, j=1,2,\dots,n$, di un vettore \mathbf{V} (punto P) di \mathbb{R}^n , si ottiene in modo immediato dall’identità vettoriale:

$$\mathbf{v} = v_1 \mathbf{a}_1 + v_2 \mathbf{a}_2 + \dots + v_n \mathbf{a}_n = \tilde{v}_1 \tilde{\mathbf{a}}_1 + \tilde{v}_2 \tilde{\mathbf{a}}_2 + \dots + \tilde{v}_n \tilde{\mathbf{a}}_n.$$

ovvero, sfruttando la notazione matriciale:

$$\mathbf{v} = \mathbf{I}_n \cdot \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = [\tilde{\mathbf{a}}_1 \quad \tilde{\mathbf{a}}_2 \quad \dots \quad \tilde{\mathbf{a}}_n] \cdot \begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_2 \\ \vdots \\ \tilde{v}_n \end{pmatrix}$$

dove \mathbf{I}_n è la matrice identità $n \times n$. Poiché la matrice le cui colonne sono costituite dai versori $\tilde{\mathbf{a}}_j, j=1,2,\dots,n$, è ortogonale, essa è invertibile e la sua inversa si ottiene semplicemente facendone la trasposta. L’espressione precedente è quindi intesa come sistema nelle incognite $\tilde{v}_j, j=1,2,\dots,n$, ed ha come soluzione:

$$\begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_2 \\ \vdots \\ \tilde{v}_n \end{pmatrix} = \begin{bmatrix} \tilde{\alpha}'_1 \\ \tilde{\alpha}'_2 \\ \vdots \\ \tilde{\alpha}'_n \end{bmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \mathbf{v}'\tilde{\alpha}_1 \\ \mathbf{v}'\tilde{\alpha}_2 \\ \vdots \\ \mathbf{v}'\tilde{\alpha}_n \end{pmatrix}$$

tenuto conto del fatto che il prodotto scalare è commutativo: $\tilde{\alpha}'_1 \mathbf{v} = \mathbf{v}'\tilde{\alpha}_1, \dots$ e così via. Risulta evidente come l'elemento j -esimo della soluzione cercata corrisponda alla proiezione ortogonale di \mathbf{V} sull'asse coordinato j -esimo.

Sia $V_{(k)}$ una varietà lineare di \mathbb{R}^n di rango pieno; si può dimostrare che ad un suo generico punto P (e relativo vettore $\boldsymbol{\eta}$ associato) si può dare una rappresentazione rispetto ad un sistema di versori ortonormali $\tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_k$, tutti appartenenti a $V_{(k)}$, cioè mediante un sistema di coordinate cartesiane ortogonali proprio della varietà.

Il legame fra le due rappresentazioni è stabilito, in analogia a quanto già detto, dall'identità vettoriale:

$$\boldsymbol{\eta} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_k \mathbf{x}_k = \tilde{\beta}_1 \tilde{\alpha}_1 + \tilde{\beta}_2 \tilde{\alpha}_2 + \dots + \tilde{\beta}_k \tilde{\alpha}_k.$$

Tale identità deve intendersi come un sistema lineare nelle incognite $\tilde{\beta}_j, j=1,2,\dots,k$, e se è risolubile (il che è vero per ipotesi) la soluzione è unica. Per trovarla basta moltiplicare ambo i membri successivamente per $\tilde{\alpha}'_1, \tilde{\alpha}'_2, \dots, \tilde{\alpha}'_k$, tenuto conto dell'ortonormalità di tali versori. Si ottiene in corrispondenza, con le notazioni già introdotte, la seguente:

$$\begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \vdots \\ \tilde{\beta}_k \end{pmatrix} = \begin{bmatrix} \tilde{\alpha}'_1 \\ \tilde{\alpha}'_2 \\ \vdots \\ \tilde{\alpha}'_k \end{bmatrix} \cdot \mathbf{x}\boldsymbol{\beta}$$

Tale relazione stabilisce il legame richiesto e dà anche le nuove componenti $\tilde{\beta}_j$ di $\boldsymbol{\eta}$ nel sistema di riferimento ortonormale istituito in $V_{(k)}$.

Sussiste, infine, il seguente teorema.

Teorema 1.2.2

- (a) Se $V_{(k)}$ è una varietà lineare di \mathbb{R}^n , $1 < k < n$, è possibile definire una base ortonormale $\tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_k$, per $V_{(k)}$, costituita da versori appartenenti a $V_{(k)}$, sia nel caso di rango pieno che non.
- (b) E' possibile, inoltre, completare la base anzidetta con altri $(n-k)$ versori $\tilde{\alpha}_{k+1}, \tilde{\alpha}_{k+2}, \dots, \tilde{\alpha}_n$, non appartenenti a $V_{(k)}$, tali che $\tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_k, \tilde{\alpha}_{k+1}, \tilde{\alpha}_{k+2}, \dots, \tilde{\alpha}_n$ costituiscano una base ortonormale per \mathbb{R}^n .

Questo teorema assicura la proprietà intuitiva che, riferita la varietà $V_{(k)}$ ad un proprio sistema di riferimento cartesiano ortogonale, caratterizzato da un numero di assi pari alla sua dimensione, ovvero k , il sistema può completarsi con assi ortogonali “esterni” alla varietà, in modo da avere un sistema di riferimento cartesiano ortogonale complessivo per \mathbb{R}^n .

1.3 Il modello di regressione lineare nei parametri

Per modello statistico **lineare nei parametri** intendiamo un modello interpretativo in base al quale il risultato y di una osservazione sia assunto essere determinazione di una variabile casuale Y del tipo:

$$Y = x_1\beta_1 + x_2\beta_2 + \dots + x_m\beta_m + E$$

dove le $x_j, j=1,2,\dots,m$ sono valori costanti e noti, i $\beta_j, j=1,2,\dots,m$ sono incogniti parametri reali, ed E è una variabile casuale (componente d'errore) di media nulla e varianza finita, che riassume la componente di variabilità accidentale non controllabile.

A tale modello corrisponde quello teorico per la media della variabile Y :

$$M(Y) = x_1\beta_1 + x_2\beta_2 + \dots + x_m\beta_m = \eta .$$

Nell'impiego pratico l'utilizzo di un modello di regressione richiede più osservazioni della variabile Y , corrispondenti a differenti livelli dei regressori x , di conseguenza il modello complessivo sarà del tipo:

$$Y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + \dots + x_{mi}\beta_m + E_i, \quad i = 1, 2, \dots, n$$

che, rappresentando secondo la notazione matriciale già introdotta le sue componenti come segue:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_n \end{pmatrix}$$
$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{m1} \\ x_{12} & x_{22} & \dots & x_{m2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \dots & x_{mn} \end{bmatrix}$$

viene ad essere esprimibile in modo più compatto come:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}.$$

A riguardo della componente d'errore vengono fatte le seguenti assunzioni:

$$M(\mathbf{E}) = \mathbf{0}, \quad \text{Var}(\mathbf{E}) = \sigma^2 \cdot \mathbf{I}_n$$

ovvero si suppone che \mathbf{E} costituisca una variabile casuale n -dimensionale con componenti di media nulla, comune varianza e fra loro non correlate.

Alla luce di questa formulazione e delle assunzioni fatte siamo in grado di esprimere il livello medio della risposta \mathbf{Y} come segue:

$$M(\mathbf{Y}) = M \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = M \begin{pmatrix} M(Y_1) \\ M(Y_2) \\ \vdots \\ M(Y_n) \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

e lo stesso siamo in grado di fare per la struttura di variabilità:

$$\begin{aligned} \boldsymbol{\Sigma}_Y &= M[(\mathbf{Y} - \boldsymbol{\eta})'(\mathbf{Y} - \boldsymbol{\eta})] = M[\mathbf{E}'\mathbf{E}] = \\ &= \begin{bmatrix} M(E_1^2) & M(E_1E_2) & \dots & M(E_1E_n) \\ M(E_2E_1) & M(E_2^2) & \dots & M(E_2E_n) \\ \vdots & \vdots & \ddots & \vdots \\ M(E_nE_1) & M(E_nE_2) & \dots & M(E_n^2) \end{bmatrix} = \\ &= \boldsymbol{\Sigma}_E = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \cdot \mathbf{I}_n \end{aligned}$$

dove Σ_Y, Σ_E , indicano rispettivamente la matrice di varianze e covarianze dei vettori casuali \mathbf{Y}, \mathbf{E} .

Portando quindi l'attenzione sulle medie, intese come funzioni dei parametri β , cioè sull'espressione

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m] \boldsymbol{\beta} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_m \mathbf{x}_m$$

è interessante notare come, alla luce di quanto già discusso, questa relazione imponga alle medie delle variabili casuali oggetti di studio di appartenere (in situazioni non degeneri) nel caso bidimensionale ad una retta, in quello tridimensionale ad una retta o ad un piano; in generale ad una entità che abbiamo chiamato **varietà lineare** e che è l'estensione della retta e del piano.

Quindi l'aver ipotizzato che le osservazioni sono tali che le rispettive medie sono tutte esprimibili mediante gli stessi parametri β fa sì che le variabili casuali in oggetto non siano più del tutto generiche, ma tali che le rispettive medie appartengano, come già detto, ad un ente matematico da noi definito come varietà lineare.

Per chiarire quanto affermato si consideri il caso semplice in cui la matrice \mathbf{X} presenti due sole colonne, in corrispondenza avremo:

$$\boldsymbol{\eta} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 = \beta_1 \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \end{pmatrix} + \beta_2 \begin{pmatrix} x_{21} \\ x_{22} \\ x_{23} \end{pmatrix}$$

dove \mathbf{x}_1 e \mathbf{x}_2 sono vettori, ad esempio di uno spazio tridimensionale, che formano le due colonne della matrice. È immediato quindi vedere a quale insieme di punti è vincolato il vettore risultante $\boldsymbol{\eta}$ al variare di β .

Nel riferimento cartesiano ortogonale tridimensionale si possono sempre rappresentare i vettori \mathbf{x}_1 e \mathbf{x}_2 , secondo le convenzioni della rappresentazione

geometrica dei vettori, applicandoli nell'origine. Essi individuano un piano che sarà degenere, riducendosi ad una retta, nel caso in cui essi siano linearmente dipendenti. Si deve quindi distinguere il caso in cui la caratteristica della matrice \mathbf{X} sia uguale, da quello in cui sia inferiore al numero dei parametri.

Nel caso in cui \mathbf{X} abbia caratteristica due si riscontra immediatamente che il vettore $\boldsymbol{\eta}$, le cui componenti sono componenti del vettore delle medie, giace necessariamente nel piano anzidetto. Infatti $\boldsymbol{\eta}$ è ottenuto come combinazione lineare dei vettori \mathbf{x}_1 , \mathbf{x}_2 e si trova pertanto sul piano da essi individuato.

Nel caso degenere, in cui le direzioni dei due vettori non sono distinte, nulla cambia concettualmente, ma il vincolo è rappresentato da una retta, poiché \mathbf{x}_1 , \mathbf{x}_2 individuano un'unica retta e quindi, in definitiva, $\boldsymbol{\eta}$ sarà vincolato ad appartenere a tale retta.

Si può studiare il problema da un altro punto di vista, ovvero scrivendo esplicitamente le componenti del vettore $\boldsymbol{\eta}$; ciò porta ad un numero di equazioni scalari pari a quello delle componenti del vettore.

Pertanto, nel caso esaminato, avremo il seguente sistema di tre equazioni:

$$\begin{cases} \eta_1 = \beta_1 x_{11} + \beta_2 x_{21} \\ \eta_2 = \beta_1 x_{12} + \beta_2 x_{22} \\ \eta_3 = \beta_1 x_{13} + \beta_2 x_{23}. \end{cases}$$

Si può ora, nell'ipotesi che le colonne siano linearmente indipendenti, risolvere il sistema rispetto alle prime due equazioni ottenendo:

$$\begin{aligned} \beta_1 &= f_1(\eta_1, \eta_2) \\ \beta_2 &= f_2(\eta_1, \eta_2) \end{aligned}$$

e, introducendo i valori così trovati nella terza espressione, si ottiene alla fine una relazione lineare in η_1, η_2, η_3 , cioè l'equazione di un piano.

Pertanto si nota che anche in base a considerazioni algebriche si giunge a stabilire il significato geometrico del modello lineare sopra discusso.

L'interpretazione geometrica può, in seconda istanza, estendersi al caso di uno spazio \mathbb{R}^n a un numero qualsiasi di dimensioni in base a quanto visto.

1.4 Stima dei parametri di un modello lineare

Il metodo di stima ottimale per il problema in esame è quello basato sul noto criterio o principio dei “minimi quadrati ordinari”.

Per affrontarlo dal punto di vista geometrico si rende necessario introdurre le rappresentazioni vettoriali di alcune quantità. Si vuole in primo luogo scrivere in forma vettoriale l'espressione, sia essa $S(\mathbf{y}, \boldsymbol{\beta})$, che deve essere minimizzata per l'applicazione del criterio dei minimi quadrati. Precisamente:

$$S(\mathbf{y}, \boldsymbol{\beta}) = \sum_{i=1}^n \left(y_i - \sum_{j=1}^m x_{ji} \beta_j \right)^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

è l'espressione da considerare. Tale espressione è costituita dalla somma di certe quantità elevate al quadrato, ciascuna delle quali si riconosce essere la differenza fra un'osservazione ed il valore assegnato dal modello alla media della corrispondente variabile casuale, espressa qui nella notazione precedentemente introdotta.

Il principio dei minimi quadrati consiste nello scegliere quei valori di $\boldsymbol{\beta}$ che minimizzano $S(\mathbf{y}, \boldsymbol{\beta})$. In dettaglio tale espressione rappresenta il modulo al quadrato del vettore

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \begin{pmatrix} y_1 - \sum_{j=1}^m x_{j1} \beta_j \\ y_2 - \sum_{j=1}^m x_{j2} \beta_j \\ \vdots \\ y_n - \sum_{j=1}^m x_{jn} \beta_j \end{pmatrix}.$$

Ciò premesso si può risolvere il problema di minimizzazione di $S(\mathbf{y}, \boldsymbol{\beta})$ tramite l'approccio differenziale. Questo afferma essere condizione necessaria e sufficiente affinché particolari valori β_j la minimizzino, che per gli stessi le derivate parziali rispetto alle incognite β_j siano tutte simultaneamente nulle, cioè risulti:

$$\frac{\partial S}{\partial \beta_h} = 0, \quad h = 1, 2, \dots, m.$$

Derivando ed uguagliando a zero le m espressioni risultanti si ottiene un sistema di equazioni, denominato **sistema delle equazioni normali**, che presenta lungo la diagonale principale della matrice associata (a meno della costante di proporzionalità $1/n$) i momenti secondi rispetto all'origine dei valori x delle variabili x_j (i regressori); al di fuori di detta diagonale figurano, a meno della stessa costante di proporzionalità, gli analoghi momenti misti fra coppie di tali variabili.

Da ciò discende che la matrice del sistema è simmetrica. Introducendo la scrittura compatta, in termini matriciali, avremo la seguente rappresentazione equivalente:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y};$$

$\mathbf{X}'\mathbf{X}$ esprime quindi la matrice del sistema delle equazioni normali.

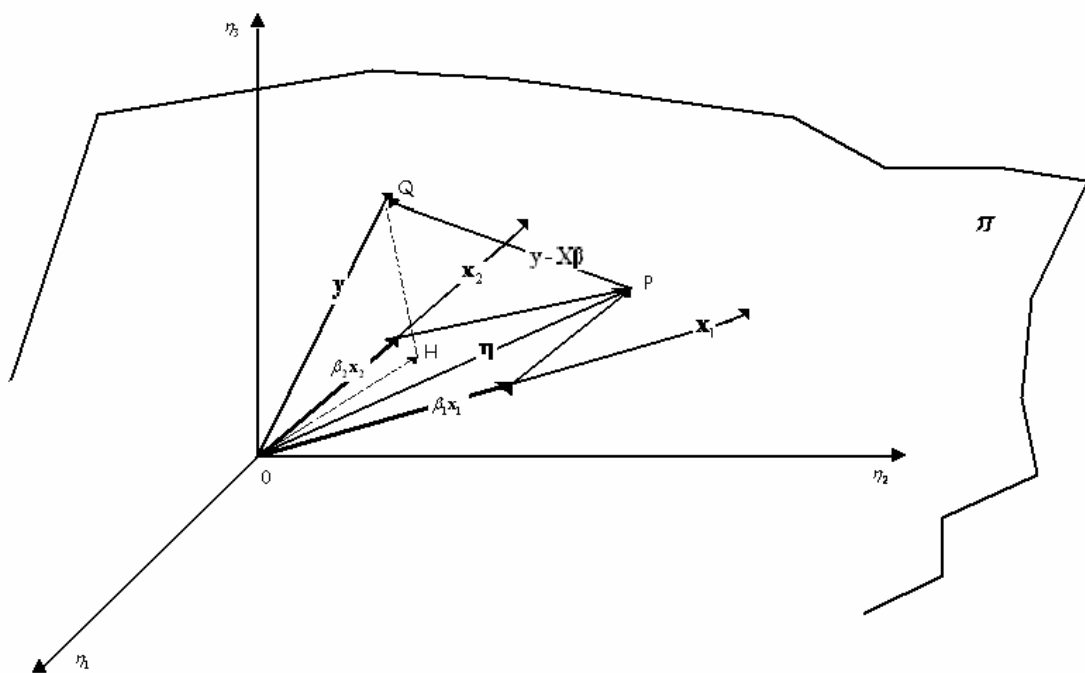
Bisogna ora stabilire se tale sistema sia o meno risolubile, potrebbe pensarsi che, in particolari condizioni, non lo sia, e non esista in tal caso alcun minimo; in secondo luogo, trovate le soluzioni, bisogna verificare se esse sono effettivamente soluzioni di minimo. Interpretaremo geometricamente entrambi questi problemi.

Si noti in primo luogo come, secondo quanto già detto, un vettore del tipo $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ appartiene ad una varietà lineare $V_{(k)}$, dove k è la dimensione di tale varietà nonché il rango della matrice \mathbf{X} (nella rappresentazione grafica che segue, con riferimento ad uno spazio \mathbb{R}^3 una varietà lineare viene rappresentata con il piano π individuato dai due vettori \mathbf{x}_1 e \mathbf{x}_2).

Si tratta quindi, assegnato il vettore delle osservazioni \mathbf{y} , di dimensione pari a quella di $\boldsymbol{\eta}$, che pertanto può essere rappresentato nello stesso sistema di riferimento, di rendere minima la distanza del “vertice” di tale vettore da un generico punto della varietà cui appartiene il vettore $\boldsymbol{\eta}$.

L'espressione $[S(\mathbf{y}, \boldsymbol{\beta})]^{1/2}$ dà infatti la distanza di \mathbf{y} da un punto qualsiasi di $V_{(k)}$, poiché è formalmente identica alla formula della distanza tra due punti dello spazio euclideo.

Con riferimento allo spazio \mathbb{R}^3 , per fissare le idee, consideriamo la figura che segue; ricordando che $\boldsymbol{\eta} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 = \mathbf{X}\boldsymbol{\beta}$, per la definizione di differenza di due vettori si ha: $Q - P = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ e si vuole rendere minimo il modulo (al quadrato) di tale vettore, dato da $S(\mathbf{y}, \boldsymbol{\beta})$.



È quindi chiaro che calando da Q la perpendicolare al piano π si ha la risposta al primo quesito: i valori $\hat{\beta}_1$ e $\hat{\beta}_2$ corrispondenti ad H , infatti, lo risolvono, dato che qualsiasi altro punto P del piano ha certamente distanza maggiore di H da Q .

Precisamente si stabilisce l'esistenza di una soluzione per il sistema delle equazioni normali in base al ragionamento che segue. Se si riflette sul significato che hanno

nella rappresentazione geometrica i valori β_1 e β_2 si nota la completa analogia tra la scrittura

$$\boldsymbol{\eta} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2$$

e quella di un vettore \mathbf{V} attraverso i versori:

$$\mathbf{v} = v_1 \mathbf{a}_1 + v_2 \mathbf{a}_2.$$

Segue che i valori β_1 e β_2 possono interpretarsi come particolari coordinate, riferite però ad un sistema di assi obliquo, poiché, in generale, \mathbf{x}_1 e \mathbf{x}_2 non sono ortogonali, mentre tali sono per ipotesi \mathbf{a}_1 e \mathbf{a}_2 .

Pertanto, in corrispondenza ad H , che è il punto immagine della soluzione geometrica del problema, si avranno in generale certe coordinate oblique $\hat{\beta}_1$ e $\hat{\beta}_2$, costruite in modo tale che il vettore $H-O$ sia definito mediante la nota regola del parallelogramma. Visto in termini geometrici il problema porta chiaramente ad un'unica soluzione, dato che unico è il piede H della perpendicolare; questa affermazione vale nel caso esaminato nella figura poiché i vettori \mathbf{x}_1 , \mathbf{x}_2 sono stati assunti essere linearmente indipendenti, e quindi esiste effettivamente il piano π .

Se invece si suppongono \mathbf{x}_1 , \mathbf{x}_2 linearmente dipendenti, è immediato convincersi che, ferma restando l'unicità della soluzione geometrica, esistono infinite soluzioni del problema algebrico, ovvero del sistema delle equazioni normali.

Infatti se:

$$\mathbf{x}_2 = \lambda \mathbf{x}_1, \quad -\infty < \lambda < \infty$$

si ha che il vettore \mathbf{x}_2 ha la stessa direzione del vettore \mathbf{x}_1 ed il modello, in questo caso, assume la forma:

$$\boldsymbol{\eta} = \beta_1 \mathbf{x}_1 + \lambda \beta_2 \mathbf{x}_1 = (\beta_1 + \lambda \beta_2) \mathbf{x}_1,$$

che rappresenta l'equazione di una retta.

In questa situazione esiste ancora ed è unica la proiezione $H-O$ del vettore \mathbf{y} sul vettore $\boldsymbol{\eta}$ in accordo ai ragionamenti prima fatti. In corrispondenza ad H esisterà, quindi, un valore $\hat{\beta}^*$, che, però, espresso rispetto agli assi originari è del tipo:

$$\hat{\beta}^* = \hat{\beta}_1 + \lambda \hat{\beta}_2.$$

Pertanto, poiché $\hat{\beta}^*$ è espresso mediante le due variabili $\hat{\beta}_1$ e $\hat{\beta}_2$, esistono infiniti modi di ottenere uno stesso valore $\hat{\beta}^*$ mediante una coppia $\hat{\beta}_1, \hat{\beta}_2$.

Si è quindi accertato che in questo caso esistono infinite soluzioni di minimo nei parametri originali, e quindi del problema che ci siamo posti.

Concludendo si può affermare che il sistema ha, in generale, almeno una soluzione, e può averne, in casi particolari, infinite.

L'estensione formale ad $\mathbb{R}^n, n > 3$, delle nozioni e proprietà che caratterizzano lo spazio intuitivo \mathbb{R}^3 , in particolare delle nozioni di ortogonalità e proiezione ortogonale, consentono di rendere più rigorose e generali le precedenti conclusioni intuitive. Ciò è stabilito dal seguente risultato.

Teorema 1.2.3

Si consideri un modello lineare nei parametri ed il corrispondente sistema di equazioni normali.

1. Esiste sempre una soluzione $\hat{\beta}_j, j=1,2,\dots,m$ di tale sistema, che minimizza l'espressione $S(\mathbf{y}, \boldsymbol{\beta})$. Se $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ è la proiezione ortogonale del vettore \mathbf{Y} , riassuntivo delle osservazioni, sulla varietà lineare $V_{(k)}$ definita dal vettore delle medie $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, proiezione la cui esistenza è assicurata dal Teorema 1.2.1, le componenti $\hat{\beta}_j, j=1,2,\dots,m$ di un vettore $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m)'$: $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ costituiscono tale soluzione.

2. Ogni soluzione $\hat{\beta}_j, j=1,2,\dots,m$ del sistema delle equazioni normali minimizza l'espressione $S(\mathbf{y}, \boldsymbol{\beta})$. Essa stabilisce infatti una (la) rappresentazione $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m)'$, della proiezione ortogonale di \mathbf{y} su $V_{(k)}$.

Dimostrazione

1. Ricordando che:

$$S(\mathbf{y}, \boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = (\mathbf{y} - \boldsymbol{\eta})' (\mathbf{y} - \boldsymbol{\eta})$$

si vuole in primo luogo dimostrare che la proiezione di \mathbf{y} sulla varietà è caratterizzata da valori $\hat{\beta}_j$ che soddisfano al sistema delle equazioni normali. Sia $\hat{\boldsymbol{\eta}}$ detta proiezione, che sappiamo esistere in virtù del Teorema 1.2.1. Si aggiunga e si tolga nella precedente espressione $\hat{\boldsymbol{\eta}}$ come sotto indicato:

$$\begin{aligned} S(\mathbf{y}, \boldsymbol{\beta}) &= [(\mathbf{y} - \hat{\boldsymbol{\eta}}) + (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})]' [(\mathbf{y} - \hat{\boldsymbol{\eta}}) + (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})] = \\ &= (\mathbf{y} - \hat{\boldsymbol{\eta}})' (\mathbf{y} - \hat{\boldsymbol{\eta}}) + (\mathbf{y} - \hat{\boldsymbol{\eta}})' (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) + \\ &+ (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})' (\mathbf{y} - \hat{\boldsymbol{\eta}}) + (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})' (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = \\ &= (\mathbf{y} - \hat{\boldsymbol{\eta}})' (\mathbf{y} - \hat{\boldsymbol{\eta}}) + (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})' (\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \end{aligned}$$

tenuto conto dell'ortogonalità dei vettori $(\mathbf{y} - \hat{\boldsymbol{\eta}})$ e $(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$, dovuta al fatto che $\hat{\boldsymbol{\eta}}$ è la proiezione sulla varietà $V_{(k)}$ definita da $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, ed $\boldsymbol{\eta}, \hat{\boldsymbol{\eta}}$ (e quindi $(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta})$) appartengono a $V_{(k)}$. Si vede pertanto che $S(\mathbf{y}, \boldsymbol{\beta})$ diventa minimo se e solo se $\boldsymbol{\eta}$ coincide con $\hat{\boldsymbol{\eta}}$, essendo \mathbf{y} un vettore prefissato.

In definitiva la scelta di $\hat{\boldsymbol{\eta}}$ e, quindi, di valori $\beta_j : \hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, che certo esistono poiché $\hat{\boldsymbol{\eta}} \in V_{(k)}$, porta ad una soluzione del problema di minimo in oggetto, poiché

minimizza la quantità d'interesse. È ora immediato verificare che i valori $\hat{\beta}_j$ anzidetti, definiti come “coordinate” della proiezione $\hat{\boldsymbol{\eta}}$ di \mathbf{y} su $V_{(k)}$, soddisfano al sistema delle equazioni normali; si consideri la differenza $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$: poiché $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ è la proiezione di \mathbf{y} su $V_{(k)}$, essa è ortogonale ad ogni vettore della varietà. Fra tutti i vettori di quest'ultima esiste anche il vettore corrispondente alla prima colonna di \mathbf{X} , alla seconda, alla terza, e così via. Si consideri al riguardo la scomposizione:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m],$$

ora, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ appartengono sicuramente alla varietà, poiché ognuno di essi è del tipo $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, e quindi la proprietà di ortogonalità dovrà verificarsi anche per ognuno di questi vettori, cioè:

$$\begin{cases} \mathbf{x}'_1 (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0 \\ \mathbf{x}'_2 (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0 \\ \dots \\ \mathbf{x}'_m (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0 \end{cases}$$

In forma più compatta in termini matriciali si può scrivere:

$$\begin{bmatrix} \mathbf{x}'_1 \\ \mathbf{x}'_2 \\ \vdots \\ \mathbf{x}'_m \end{bmatrix} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0,$$

ovvero:

$$\mathbf{X}' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0$$

da cui

$$\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0$$

e quindi infine:

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}},$$

che è proprio l'equazione che si voleva verificare, in quanto rappresenta compattamente il sistema delle equazioni normali.

Si è così dimostrato che esiste sempre una soluzione delle equazioni normali che corrisponde ad una soluzione di minimo del problema.

Il fatto che esista sempre è legato al fatto che la proiezione esiste sempre, e quindi esistono sempre dei valori $\hat{\beta}_j$ che soddisfano alle equazioni normali e, per quanto si è detto, corrispondono ad una soluzione di minimo.

2. Il secondo comma del teorema afferma che ogni soluzione del sistema delle equazioni normali è soluzione di minimo. Per dimostrarlo basta ripercorrere a ritroso la successione di passi che ha portato alla dimostrazione del primo comma. Si parte quindi dalle equazioni normali e si fissa l'attenzione su una generica soluzione, supponendo, quindi, di aver trovato un vettore $\hat{\boldsymbol{\beta}}$ soddisfacente al sopraccitato sistema. È allora sempre possibile esprimere il sistema nella forma equivalente

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0.$$

Da questa posizione si torna alla

$$\begin{cases} \mathbf{x}'_1(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0 \\ \mathbf{x}'_2(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0 \\ \dots \\ \mathbf{x}'_m(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0 \end{cases},$$

la quale afferma che il vettore che si è costruito come differenza dei vettori \mathbf{y} e $\mathbf{X}\hat{\boldsymbol{\beta}}$ è ortogonale al primo vettore di definizione della varietà, al secondo, al terzo, e così via fino all' m -esimo. Si può allora concludere che è anche ortogonale all'intera varietà, in quanto il generico vettore $\boldsymbol{\eta}$ della varietà può essere espresso come:

$$\boldsymbol{\eta} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_m \mathbf{x}_m.$$

Se si considera ora il prodotto scalare di $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, che soddisfa ciascuna delle equazioni del precedente sistema, per $\boldsymbol{\eta}$ ci si convince immediatamente che soddisfa anche l'equazione $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \boldsymbol{\eta} = 0$; infatti detta relazione può risciversi come:

$$(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_m \mathbf{x}_m) = 0,$$

certo soddisfatta poiché il prodotto scalare $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{x}_1$ è nullo, in base a quanto già detto, e così via per gli altri prodotti, per qualsiasi β_j . Ciò è intuitivo: infatti se il vettore $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ è ortogonale a tutti i vettori $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ che definiscono la varietà $V_{(k)}$, si comprende che sarà pure ortogonale alla varietà.

In definitiva si è stabilito che $\mathbf{X}\hat{\boldsymbol{\beta}}$ è la proiezione di \mathbf{y} sulla varietà, qualunque sia la soluzione del sistema delle equazioni normali e, quindi, assicura il valore minimo di $S(\mathbf{y}, \boldsymbol{\beta})$, che è quanto si voleva dimostrare. \square

Conviene sottolineare ancora al riguardo del precedente teorema che, pur essendo la proiezione di \mathbf{y} su $V_{(k)}$ unica, le soluzioni del sistema delle equazioni normali possono essere infinite. Ciò significa che possono esistere infinite rappresentazioni algebriche del vettore proiezione $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ mediante coordinate del tipo $\hat{\beta}_j, j = 1, 2, \dots, m$. Se si rivedono le premesse intuitive al Teorema 1.2.3, presentate anche graficamente per ovvia necessità nel caso dello spazio \mathbb{R}^3 , che si estendono in modo immediato al caso generale, esse portano alla conclusione che la soluzione del sistema delle equazioni normali non è unica se e solo se il modello lineare non è di rango pieno, vale a dire, alcune colonne della matrice \mathbf{X} possono esprimersi come combinazioni lineari delle restanti.

1.5 Approssimazione mediante linearizzazione

Supponiamo ora che la relazione intercorrente fra le variabili in esame, suggerita da una opportuna rappresentazione grafica dei dati o dalle conoscenze in possesso dello sperimentatore, sia da ritenersi non appartenente alla classe dei modelli di regressione lineari delineati in precedenza, ovvero che, secondo la definizione già data, gli incogniti parametri compaiano nel modello in modo non lineare. Un esempio molto semplice è dato dal modello $y = x^{\vartheta} + \varepsilon$, dove ϑ rappresenta l'incognito parametro da stimare e le condizioni sulla componente di errore restano le stesse già presentate.

Così come per i modelli lineari possiamo provare a calcolare il valore del coefficiente tramite il criterio dei LS; questa strada non porta però ad una espressione esplicita per la stima del parametro, anzi, le equazioni risultanti conducono ad un sistema non lineare, risolubile solo tramite procedure iterative che prendano l'avvio da un qualche valore (arbitrario) per ϑ .

Altro punto fondamentale è che lo stimatore LS per un modello non lineare non possiede le stesse proprietà di quello ottenuto applicando il criterio ad un modello lineare; solo asintoticamente, cioè al tendere della dimensione campionaria verso

l'infinito, possiamo sperare di ottenerle. In sostanza in presenza di campioni finiti, ovvero la situazione davanti a cui si trova ogni sperimentatore, se pure la variabile dipendente dovesse risultare normale in virtù delle assunzioni sugli errori, la non linearità porterebbe a non garantire l'ottenimento di stimatori normalmente distribuiti, in quanto combinazioni non lineari dei valori della y , né la non distorsione, né tantomeno la proprietà di minima varianza. Dunque, a differenza delle stime LS per un modello lineare, quelle per un modello non lineare possiedono proprietà essenzialmente ignote per campioni di dimensione finita; in generale non è comunque possibile presentare alcuna indicazione di carattere assoluto riguardo a quanto dovrebbe essere ampio il campione per l'avvicinamento di queste proprietà asintotiche. Vi sono alcuni modelli non lineari per i quali tali proprietà sono una buona approssimazione anche per campioni di piccole dimensioni, così come ne esistono altri per cui anche ciò che in termini pratici può essere considerato un campione molto grande non permette di fornire che povere approssimazioni delle proprietà asintotiche.

La maggior parte degli algoritmi disponibili per il calcolo delle stime $\hat{\boldsymbol{\theta}}$ secondo i LS e gran parte delle tecniche inferenziali riguardanti i modelli non lineari sono basati su di una approssimazione lineare locale del modello. Nell'intorno di un fissato valore parametrico $\boldsymbol{\theta}_0$ la funzione del modello è approssimata da

$$f(\mathbf{x}, \boldsymbol{\theta}) \cong f(\mathbf{x}, \boldsymbol{\theta}_0) + \sum_{i=1}^p (\mathcal{G}_i - \mathcal{G}_{i,0}) v_i(\mathbf{x})$$

dove

$$v_i(\mathbf{x}) = \partial f(\mathbf{x}, \boldsymbol{\theta}_0) / \partial \mathcal{G}_i \Big|_{\boldsymbol{\theta}_0}.$$

Raggruppando le componenti nei vettori $\boldsymbol{\eta}(\boldsymbol{\theta})$, $\boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$ e \mathbf{v}_i , dove

$$\begin{aligned}\mathbf{v}_i &= (v_i(\mathbf{x}_1), v_i(\mathbf{x}_2), \dots, v_i(\mathbf{x}_n))' = \\ &= \partial f \boldsymbol{\eta} / \partial \mathcal{G}_i \big|_{\mathcal{G}_0}, \quad i = 1, 2, \dots, p,\end{aligned}$$

otteniamo

$$\boldsymbol{\eta}(\boldsymbol{\theta}) \cong \boldsymbol{\eta}(\boldsymbol{\theta}_0) + \sum_{i=1}^p (\mathcal{G}_i - \mathcal{G}_{i,0}) \mathbf{v}_i.$$

L'effetto di tale approssimazione è di sostituire il luogo delle soluzioni con il suo piano tangente in $\boldsymbol{\eta}(\boldsymbol{\theta}_0)$, e simultaneamente di imporre un sistema di coordinate uniformi su detto piano. Entrambe l'efficacia degli algoritmi di stima secondo LS e la validità dell'inferenza applicata ai parametri di un modello non lineare saranno fortemente influenzate dalla vicinanza dell'approssimazione lineare al modello.

1.6 Intenti del lavoro

Ci chiediamo quindi come poter analizzare in modo più preciso e rigoroso la perdita di validità degli stimatori dovuta alla non linearità delle relazioni funzionali definitorie dello stesso modello. Facciamo dunque appello alla Geometria Differenziale, branca della matematica dedicata allo studio delle proprietà degli spazi tramite lo strumento, appunto, del calcolo differenziale; è possibile infatti leggere in chiave geometrica un modello interpretativo (si è visto nel caso semplice del modello lineare), come uno spazio metrico di dimensione pari al numero dei parametri, ed *immerso* in uno spazio di dimensione strettamente maggiore pari al numero delle osservazioni. Come è noto è infatti necessario disporre di un numero di osservazioni campionarie strettamente maggiore del numero dei parametri al fine di poterli stimare ed avere i “gradi di libertà” sufficienti per valutarne la variabilità ed applicare le consuete procedure inferenziali. Come già affermato, lavorando con un modello lineare l'interpretazione geometrica porta specificamente ad un

iperpiano, nel caso invece di un generico modello non lineare nei parametri avremo una cosiddetta *varietà riemanniana*, che verrà definita nel secondo capitolo (in cui si renderà necessario dare un'idea dei concetti di base del calcolo differenziale assoluto, strumento per la trattazione che ci proponiamo di intraprendere) e si mostrerà un semplice esempio di tale interpretazione. Si definirà dunque il concetto di curvatura nel terzo capitolo, e verranno introdotte le misure note dalla Geometria Differenziale, con particolare riguardo alla differenza tra le nozioni di curvatura *estrinseca* ed *intrinseca*. Nel quarto capitolo, infine, si introdurranno le misure di curvatura proposte da Bates e Watts per valutare nello specifico la lontananza dalla linearità di un modello statistico, che distinguono tra una nozione di curvatura intrinseca nel modello e di una indotta dalla parametrizzazione scelta, e si procederà, forti degli strumenti presentati nei capitoli precedenti, ad una analisi delle proprietà degli stimatori, ottenuti tramite il principio dei minimi quadrati, nell'ottica di valutare quanto la lontananza dalla linearità influisca negativamente sulla correttezza e sull'efficienza, introducendo inoltre la misura di distorsione proposta da Box; si proporranno infine alcune analisi basate su risultati ottenuti tramite simulazioni.

CAPITOLO II

Breve introduzione al calcolo differenziale assoluto (calcolo tensoriale)

I tensori sono enti matematici che possono essere visti come generalizzazioni di scalari e vettori, e che ricoprono un ruolo fondamentale in scienze come la Fisica; molte grandezze fisiche sono infatti normalmente rappresentate da scalari e vettori: le forze sono vettori, il campo elettrico è un vettore, la temperatura ed il potenziale elettrico sono scalari, e così via. Come noto, uno scalare è rappresentato da una sola componente, i vettori dalle loro componenti rispetto ad una base che può essere n -dimensionale, a seconda del numero di dimensioni dello spazio cui si riferiscono; i tensori sono oggetti, o per meglio dire sistemi, formati da più insiemi di componenti rispetto allo spazio di riferimento.

2.1 Rappresentazione cartesiana dei vettori

Fissiamo nello spazio (e quindi a 3 dimensioni) un riferimento cartesiano ortogonale, un vettore sarà allora individuato dalle sue tre componenti cartesiane, ciascuna relativa ad uno degli assi. Indicati con x_1, x_2, x_3 gli assi, ci riferiremo alle componenti cartesiane ortogonali di un vettore \mathbf{v} con

$$v_1 \quad v_2 \quad v_3.$$

In modo più compatto, indicando gli assi con x_i o x^i ($i=1,2,3$) (e dove la i ad apice non indica un elevamento a potenza), le componenti si indicheranno con v_j o v^j ($j=1,2,3$).

Passiamo ora dalla terna cartesiana ortogonale x_i ad un'altra terna sempre cartesiana ortogonale \bar{x}_i . Le componenti cartesiane v_j del vettore \mathbf{v} rispetto alla

terna x_i si muteranno nelle componenti cartesiane \bar{v}_k di \mathbf{v} secondo la terna \bar{x}_i .
Indicati con

$$\begin{array}{ccc} \alpha_1^1 & \alpha_1^2 & \alpha_1^3 \\ \alpha_2^1 & \alpha_2^2 & \alpha_2^3 \\ \alpha_3^1 & \alpha_3^2 & \alpha_3^3 \end{array}$$

i coseni direttori di una terna rispetto all'altra, e precisamente α_k^j il coseno dell'angolo che l'asse x_j forma con l'asse \bar{x}_k , possiamo affermare, dall'algebra lineare, che il legame fra le nuove e le vecchie componenti dello stesso vettore \mathbf{v} è:

$$(2.1.1) \quad \bar{v}_k = \sum_{j=1}^3 \alpha_k^j v_j \quad (k=1,2,3).$$

Le tre componenti v_1, v_2, v_3 , relative ad una terna cartesiana prefissata, insieme alla relazione (2.1.1) che esplicita come esse si trasformino al cambiare della terna di riferimento, caratterizzano completamente il vettore \mathbf{v} e ne possono costituire una definizione. In altre parole possiamo definire il vettore nello spazio tridimensionale come quell'ente individuato, in un particolare riferimento, da tre scalari (componenti) che si trasformano secondo la legge (2.1.1) al mutare del riferimento.

Possiamo riscrivere l'espressione (2.1.1) secondo la comoda convenzione di Einstein (che impiegheremo d'ora in poi) sulla somma degli indici ripetuti, che ci consente di omettere il simbolo di sommatoria, a patto di sottintenderla riguardo agli indici che si ripetono nell'espressione:

$$(2.1.1') \quad \bar{v}_k = \alpha_k^j v_j \left(= \sum_{j=1}^3 \alpha_k^j v_j \right) \quad (k=1,2,3).$$

2.2 Invarianza delle relazioni vettoriali

Osserviamo che la (2.1.1) fornisce le nuove componenti \bar{v}_k ($k=1,2,3$) come funzioni lineari ed omogenee delle vecchie componenti v_j ($j=1,2,3$). Ne consegue che, se nel riferimento x_i la relazione

$$(2.2.1) \quad v_j = u_j \quad (j=1,2,3)$$

traduce l'uguaglianza fra i due vettori \mathbf{v} e \mathbf{u} , al mutare del riferimento le v_j si trasformano nelle \bar{v}_k , le u_j nelle \bar{u}_k , ed in virtù della (2.2.1) e della (2.1.1) si deduce che:

$$\bar{v}_k = \bar{u}_k \quad (k=1,2,3).$$

Tutte le relazioni vettoriali posseggono, rispetto alla trasformazione (2.1.1) lo stesso carattere invariante dell'uguaglianza fra vettori, carattere che rende tali relazioni atte alla rappresentazione di leggi geometriche, e più generalmente di leggi fisiche che presuppongono assoluto lo spazio di riferimento.

Quanto detto per i vettori nell'ordinario spazio tridimensionale vale anche per i vettori in un piano (due dimensioni) e più in generale per i vettori di un generico spazio euclideo ad n dimensioni. Basta rappresentare ogni vettore di questo spazio con n componenti relative ad una n -upla di assi cartesiani ortogonali, far variare gli indici da 1 a n ed estendere da 1 a n le sommatorie.

2.3 Rappresentazione cartesiana dei tensori doppi

In un prefissato riferimento cartesiano, costituito dai tre assi ortogonali x_1, x_2, x_3 , consideriamo ora tre vettori

$$\mathbf{T}_1 \quad \mathbf{T}_2 \quad \mathbf{T}_3,$$

corrispondenti ai tre assi precedenti. In forma più compatta indichiamo con \mathbf{T}_i o \mathbf{T}^i ($i=1,2,3$), il vettore corrispondente all'asse x_i . Ognuno di questi vettori potrà essere rappresentato, nel riferimento precedente, dalle sue tre componenti cartesiane: indicando con T_{ij} (o T_i^j , o T_j^i , o T^{ij}) le componenti secondo l'asse x_j del vettore \mathbf{T}_i corrispondente all'asse x_i , i tre vettori precedenti saranno rappresentati dai seguenti nove scalari:

$$\begin{array}{ccc} T_{11} & T_{21} & T_{31} \\ T_{12} & T_{22} & T_{32} \\ T_{13} & T_{23} & T_{33}. \end{array}$$

Sostituiamo, per una lettura più immediata, alla precedente formulazione la seguente scrittura:

$$T_{ij} \quad (i, j = 1, 2, 3).$$

Passiamo ora dalla terna cartesiana ortogonale x_1, x_2, x_3 ad un'altra terna, sempre cartesiana ortogonale, $\bar{x}_1, \bar{x}_2, \bar{x}_3$. Se i tre vettori \mathbf{T}_i sono collegati alla terna di riferimento come lo sono le componenti di un vettore, essi si trasformano, al mutare della terna, con la seguente legge, analoga alla (2.1.1), e sempre scritta secondo la convenzione sugli indici ripetuti:

$$(2.3.1) \quad \bar{\mathbf{T}}_h = \alpha_h^i \mathbf{T}_i \quad (h = 1, 2, 3).$$

Nella (2.3.1) \mathbf{T}_i ($i=1,2,3$) sono i vettori corrispondenti alla vecchia terna, $\bar{\mathbf{T}}_h$ ($h=1,2,3$) quelli relativi alla nuova, e α_h^i ($i, h=1,2,3$) i coseni direttori dell'una rispetto all'altra.

Introducendo in luogo dei tre vettori \mathbf{T}_i le loro nove componenti T_{ij} rispetto alla terna originaria, e in luogo dei tre vettori $\bar{\mathbf{T}}_h$ le loro nove componenti \bar{T}_{hk} rispetto alla nuova terna, dalla (2.3.1) e dalla (2.1.1) si deduce:

$$(2.3.2) \quad \bar{T}_{hk} = \alpha_h^i \alpha_k^j T_{ij} \quad (h, k = 1, 2, 3)$$

dove ancora una volta è stata impiegata la convenzione sugli indici ripetuti, in questo caso in particolare sono due gli indici che si ripetono, ovvero i e j , si tratterà dunque di una doppia sommatoria in i e j . Per quanto riguarda gli indici che scompaiono (ovvero quelli rispetto ai quali si effettua la somma) si dice che essi vengono saturati.

Così come l'assegnare tre componenti di un vettore rispetto ad una terna prefissata, e il dare la legge (2.1.1) con la quale si trasformano al cambiare della terna di riferimento caratterizza e definisce un vettore, l'assegnare i tre vettori \mathbf{T}_i rispetto ad una terna prefissata, e il dare la legge (2.3.2) con la quale essi si trasformano al cambiare della terna caratterizza completamente e definisce un ente che si dice **tensore doppio** o **di ordine 2** . Tale oggetto matematico può essere definito analogamente dando, rispetto ad una terna prefissata, i nove scalari che rappresentano le componenti dei tre vettori e la legge (2.3.2) con la quale essi si trasformano al mutare della terna. Sottintendendo quest'ultima, un tensore è individuato dalle sue componenti T_{ij} . Per questo motivo il tensore considerato è detto talvolta semplicemente tensore T_{ij} .

Un semplice esempio di tensore doppio può essere presentato considerando una quadrica a centro. Rispetto ad una terna cartesiana ortogonale x^1, x^2, x^3 (dove l'apice non indica una potenza, ma è il valore dell'indice $i=1,2,3$), avente l'origine nel centro, l'equazione della quadrica si può scrivere come segue:

$$(2.3.3) \quad T_{ij} x^i x^j = 1.$$

Fissata la terna i nove coefficienti T_{ij} ($i, j = 1, 2, 3$) individuano la quadrica.

Passiamo ora dalla terna x^1, x^2, x^3 alla terna $\bar{x}^1, \bar{x}^2, \bar{x}^3$, anch'essa cartesiana ortogonale ed avente la medesima origine della prima. Se α_h^i ($i, h = 1, 2, 3$) sono i coseni direttori di una terna rispetto all'altra, la Geometria Analitica insegna che

$$x^i = \alpha_h^i \bar{x}^h, \quad x^j = \alpha_k^j \bar{x}^k.$$

La (2.3.3) diviene pertanto:

$$T_{ij} \alpha_h^i \alpha_k^j \bar{x}^h \bar{x}^k = 1,$$

ovvero

$$(2.3.3') \quad \bar{T}_{hk} \bar{x}^h \bar{x}^k = 1,$$

se

$$\bar{T}_{hk} = \alpha_h^i \alpha_k^j T_{ij}.$$

Quest'ultima relazione coincide con la (2.3.2), e mostra che *i nove coefficienti di una quadrica a centro costituiscono un tensore doppio*.

Si noti che, mentre i coefficienti della quadrica mutano al mutare della terna di riferimento conformemente alla (2.3.2), la legge geometrica (2.3.3) rimane invariata.

Quanto detto per i tensori doppi nell'ordinario spazio tridimensionale può affermarsi analogamente per i tensori doppi nel piano e per i tensori doppi in un generico spazio euclideo ad n dimensioni, che indicheremo d'ora in poi con $E_{(n)}$: nel primo caso basta riferirsi a coppie di assi cartesiani ortogonali, a coppie di vettori, a quaterne di scalari, facendo cioè variare gli indici da 1 a 2, anziché da 1 a 3; nel

secondo caso basta riferirsi a n-uple di assi cartesiani ortogonali, a n-uple di vettore, ad insiemi di n^2 scalari, facendo cioè variare gli indici da 1 a n, anziché da 1 a 3.

2.4 Rappresentazione cartesiana dei tensori multipli

In un prefissato sistema di riferimento cartesiano ortogonale x_1, x_2, x_3 consideriamo tre tensori doppi, rappresentati da nove vettori \mathbf{T}_{ij} ($i, j = 1, 2, 3$), e quindi da ventisette scalari T_{ijr} ($i, j, r = 1, 2, 3$). Siano detti tensori legati alla terna di riferimento come lo sono le componenti di un vettore, e precisamente: nel passaggio dalla terna cartesiana originaria alla terna, pure cartesiana, $\bar{x}_1, \bar{x}_2, \bar{x}_3$, i nove vettori \mathbf{T}_{ij} ($i, j = 1, 2, 3$) si trasformino con la seguente legge, analoga alla (2.3.2):

$$\bar{\mathbf{T}}_{hk} = \alpha_h^i \alpha_k^j \mathbf{T}_{ij},$$

e, conseguentemente (in virtù della (2.1.1)), i ventisette scalari T_{ijr} ($i, j, r = 1, 2, 3$) si trasformino con la seguente legge:

$$(2.4.1) \quad \bar{T}_{hkp} = T_{ijr} \alpha_h^i \alpha_k^j \alpha_p^r.$$

L'ente testé definito è un **tensore triplo o di ordine 3**.

Analogamente un **tensore quadruplo** è rappresentabile, in un prefissato sistema di riferimento cartesiano, da tre tensori tripli, ossia da nove tensori doppi, ventisette vettori, ottantuno scalari. Nel passaggio dalla terna cartesiana ortogonale x_1, x_2, x_3 alla terna anch'essa cartesiana ortogonale $\bar{x}_1, \bar{x}_2, \bar{x}_3$, i ventisette vettori \mathbf{T}_{ijr} ($i, j, r = 1, 2, 3$) si trasformano secondo la seguente legge, analoga alla (2.4.1):

$$\bar{\mathbf{T}}_{hkp} = \mathbf{T}_{ijr} \alpha_h^i \alpha_k^j \alpha_p^r,$$

e gli ottantuno scalari T_{ijrs} ($i, j, r, s = 1, 2, 3$) con la seguente:

$$(2.4.2) \quad \bar{T}_{hkpq} = T_{ijrs} \alpha_h^i \alpha_k^j \alpha_p^r \alpha_q^s.$$

Tramite il procedimento ricorrente esposto si possono definire in generale **tensori di ordine m**, mediante 3 tensori di ordine $m-1$, 3^2 tensori di ordine $m-2, \dots, 3^{m-1}$ vettori, 3^m scalari. Nel passaggio dalla terna x_i alla terna \bar{x}_i i 3^m scalari si trasformano con una legge che generalizza ovviamente la (2.4.2).

I vettori potranno quindi essere visti come tensori di ordine 1, mentre gli scalari come tensori di ordine 0.

Come per i vettori ed i tensori doppi, può trasportarsi quanto detto nel piano ed in un generico spazio euclideo ad n dimensioni riguardo ai tensori multipli nell'ordinario spazio tridimensionale: basta a tal fine far variare gli indici, invece che da 1 a 3, da 1 a 2 nel primo caso, da 1 a n nel secondo. Chiaramente in seguito a questo, in termini generali, un tensore di ordine m in uno spazio ad n dimensioni è rappresentato da n^m scalari, n^{m-1} vettori, n^{m-2} tensori doppi, e così via.

Come già detto in inizio di discorso, i tensori sono dunque enti matematici, o per meglio dire sistemi, formati da più insiemi di componenti rispetto allo spazio cui si riferiscono. Occorre però precisare che un sistema qualunque non è obbligatoriamente un tensore, affinché lo sia è necessario che valgano alcune condizioni restrittive circa il cambiamento di questi oggetti in una trasformazione di coordinate, come già visto. I tensori sono quindi sistemi che variano secondo certe leggi quando vi è una trasformazione di coordinate. Ciò è di fondamentale importanza ad esempio per la fisica. Tenuto conto che un tensore nullo rimane tale anche dopo una qualunque trasformazione di coordinate, se le leggi della fisica fossero espresse dalla formula generale $T=0$, ove T è un tensore, allora quelle leggi sarebbero del tutto **invarianti** rispetto a qualunque trasformazione di coordinate ortogonali. Questo è, in sintesi, il principio su cui si poggia la **relatività generale**. Siccome non vi è ragione di pensare che vi siano sistemi di riferimento privilegiati

in cui le leggi della fisica abbiamo una formulazione particolare, si presuppone che il principio di relatività generale debba essere vero. Le leggi della fisica devono essere le stesse rispetto a qualunque sistema di riferimento, per cui le leggi della fisica devono, per soddisfare questo principio, essere espresse in forma tensoriale.

L'introduzione dell'ente matematico tensore è stata resa necessaria dal fatto che vi sono grandezze fisiche che non sono esprimibili soltanto in termini di scalari e vettori; se si vuole descrivere, ad esempio, la deformazione a cui è sottoposto un materiale, le grandezze vettoriali non bastano, occorrono i tensori. Il nome deriva proprio dall'impiego di questi oggetti nello studio sullo stato di *tensione* o di pressione in un punto in un corpo continuo.

2.5 Richiami di Geometria Differenziale: spazi euclidei e varietà curve

In matematica, uno **spazio euclideo** è una generalizzazione degli spazi a due e a tre dimensioni studiati da Euclide. Lo spazio euclideo è il prototipo di spazio vettoriale reale di dimensione qualsiasi (ma finita), in cui è possibile definire le nozioni di distanza, lunghezza ed angolo. Dunque lo spazio euclideo è più che un semplice spazio vettoriale. Per ottenere la cosiddetta *geometria euclidea* si deve poter parlare della distanza fra punti, e di angoli fra rette o vettori.

Il modo naturale per fare questo è l'introduzione di quello che viene chiamato *prodotto scalare* su \mathbb{R}^n definito come segue:

$$\mathbf{v} \times \mathbf{u} = v_i u_i = v_1 u_1 + v_2 u_2 + \dots + v_n u_n,$$

ove \mathbf{v} e \mathbf{u} sono vettori di \mathbb{R}^n , e si è impiegata la convenzione sugli indici ripetuti omettendo la sommatoria. Leggendo tale operazione in termini tensoriali possiamo dire che l'indice i si satura, dando luogo ad uno scalare (un ente senza indice). Questo prodotto permette la definizione di (lunghezza) di un vettore \mathbf{v} nel seguente modo:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \times \mathbf{v}},$$

quantità questa che prende il nome di **norma euclidea** di un vettore su \mathbb{R}^n . L'angolo (interno) \mathcal{G} fra due vettori \mathbf{v} e \mathbf{u} viene definito come

$$\mathcal{G} = \arccos\left(\frac{\mathbf{v} \times \mathbf{u}}{\|\mathbf{v}\| \cdot \|\mathbf{u}\|}\right)$$

e dove l'argomento della funzione inversa del coseno è in modulo minore o al più uguale ad uno, in virtù della disuguaglianza di Cauchy-Schwartz.

Infine, è possibile sfruttare la norma per definire una funzione di distanza su \mathbb{R}^n nel seguente modo:

$$d(\mathbf{v}, \mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|,$$

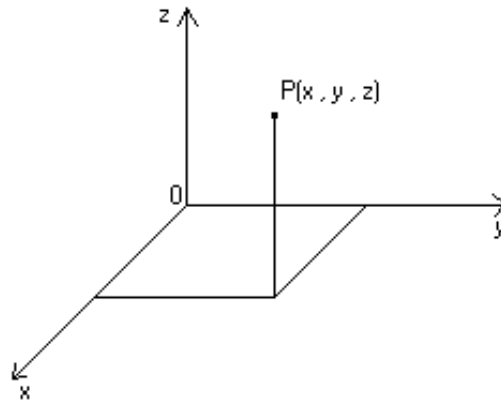
la forma di tale funzione di distanza è basata sul teorema di Pitagora, ed è chiamata **metrica euclidea**. Uno spazio ove sia possibile definire tali nozioni è quindi detto **spazio euclideo**.

Lo spazio euclideo tridimensionale è lo spazio della nostra esperienza quotidiana. Il suo nome deriva, come già premesso, dal matematico Euclide il quale, nel 300 a.C., per primo ne studiò a fondo proprietà e caratteristiche e le tramandò ai posteri in un trattato che costituisce il primo esempio conosciuto di teoria matematica impostata su basi logico-assiomatico-deduttive.

Nello spazio tridimensionale possono “disegnarsi” enti matematici (figure) a 0, 1 o 2 dimensioni; gli enti a zero dimensioni sono punti, ad una dimensione sono linee, a due superfici. Le figure dello spazio euclideo soddisfano numerose proprietà che caratterizzano lo spazio stesso. Alcune di particolare importanza sono le seguenti:

1. da un punto esterno ad una retta data passa una ed una sola retta parallela ad essa
2. per i triangoli rettangoli vale il teorema di Pitagora, ovvero il quadrato costruito sull'ipotenusa eguaglia la somma dei quadrati costruiti sui cateti
3. la somma degli angoli interni di un triangolo eguaglia l'angolo piatto (180°).

Nello spazio euclideo tridimensionale possiamo costruire un **sistema di riferimento cartesiano ortogonale** nel modo abituale ed assegnare così ad ogni punto P dello spazio una terna ordinata di numeri, le sue coordinate (x, y, z) :



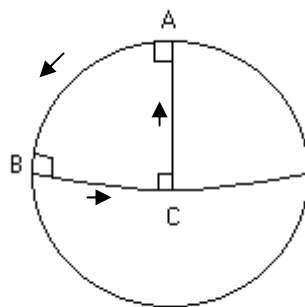
Lo spazio euclideo (detto anche **spazio piatto**) ci è così familiare che l'immaginare uno **spazio non euclideo** (detto anche **spazio curvo**), cioè per cui non valgono le nozioni presentate sopra, fu una conquista del pensiero molto recente (XIX secolo). Dapprima gli spazi non euclidei furono considerati pura astrazione matematica, poi, per opera principalmente di Einstein, divennero una necessità pratica perché egli ipotizzò, nella sua **teoria della relatività generale**, la possibilità che lo spazio fisico (quello in cui viviamo) sia euclideo solo su piccola scala, localmente, mentre su larga scala esso sia non euclideo.

Un semplice esempio di spazio non euclideo bidimensionale è quello della superficie sferica, che era stata già studiata a fondo per la necessità di disegnare carte geografiche, ma fu principalmente per opera di grandi matematici come Gauss, Christoffel, Ricci-Curbastro e Levi-Civita che questa materia fu portata ad un livello di astrazione e generalità tale da raggiungere la compiutezza teorica necessaria per essere riconosciuta come uno dei capitoli più importanti della Geometria Differenziale, ovvero il **Calcolo Differenziale Assoluto** o **Calcolo Tensoriale**.

La principale e rivoluzionaria caratteristica del calcolo tensoriale è quella di trattare spazi non euclidei di dimensione n generica con strumenti e metodologie generali che valgono, appunto, per qualsivoglia dimensione, per far sì che, all'interno di questa teoria, non sia più necessario considerare lo spazio curvo come *immerso* in uno spazio piatto di dimensione maggiore per poterne studiare le caratteristiche. Le proprietà di uno spazio curvo sono intrinseche allo spazio medesimo, a prescindere dall'immersione o meno; la distanza fra due punti in uno spazio sarà sempre la medesima, sia che si trovi in uno spazio ambiente, sia che

venga considerato a sé stante, e così le aree, i volumi, la curvatura. Il calcolo tensoriale raggiunge una così elevata generalizzazione da far sì che uno spazio piatto sia considerato solo un semplice caso particolare fra gli infiniti spazi curvi; nella fattispecie lo spazio piatto è lo spazio a **curvatura** nulla (concetto quello di curvatura centrale in questo lavoro e sul quale ritorneremo più avanti). In questa ottica lo spazio piatto non gode di alcun particolare privilegio, e le formule del calcolo tensoriale descrivono qualunque spazio curvo così come descrivono anche gli spazi euclidei (anche, e questo è un altro punto cruciale, rispetto a qualunque scelta di coordinate). In sintesi, la geometria euclidea è solo un caso particolare di una geometria più generale, quella degli spazi curvi.

Che ogni superficie curva, ed in particolare è immediato pensare alla superficie sferica, sia uno spazio non euclideo bidimensionale, significa che su di essa non valgono le regole della geometria euclidea (basta che non ne valga una sola). Come esempio consideriamo il triangolo formato sulla detta superficie sferica da un punto in movimento che parte dal polo nord A (immaginiamo che questa superficie rappresenti la superficie terrestre), arriva all'equatore in B, lo segue per un quarto della sua lunghezza fino a C ed infine ritorna al polo nord :



si tratta di un triangolo con tre angoli retti, e questo contraddice la regola euclidea riguardante la somma degli angoli interni ad un triangolo.

Procediamo dunque ad una descrizione della teoria degli spazi curvi a due dimensioni (per semplicità di trattazione) ovvero le superfici nello spazio euclideo tridimensionale, sfruttando però l'approccio del calcolo tensoriale per cui, secondo

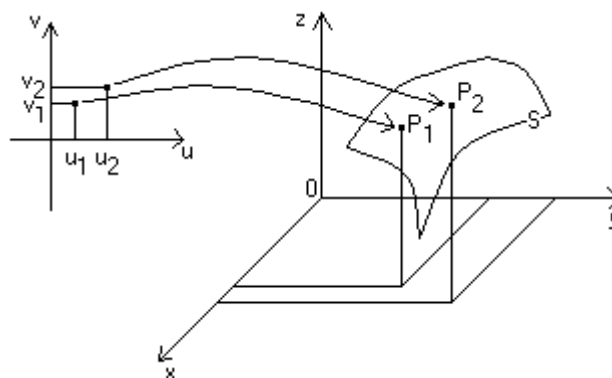
quanto finora affermato, studiando questo semplice caso immediatamente visualizzabile porremo le basi per la generalizzazione al caso degli spazi curvi n-dimensionali.

Superficie

Consideriamo una superficie S immersa in uno spazio euclideo tridimensionale in cui è definito un sistema di assi cartesiani ortogonali. Le equazioni parametriche di tale superficie sono:

$$\begin{cases} x = x(u, v) \\ y = y(u, v) \\ z = z(u, v) \end{cases}$$

dove i parametri (variabili reali indipendenti) u e v possono assumere valori entro un dominio prefissato. Il significato di queste equazioni è chiaro: al variare di u e v vengono individuati nello spazio infiniti punti di coordinate (x, y, z) ottenuti sostituendo detti valori nelle equazioni parametriche. Questi punti determinano la superficie in esame come indicato nel grafico:



Siccome le coordinate dei punti della superficie dipendono da due parametri, le equazioni sopra indicate si chiamano **equazioni parametriche della superficie**.

Bisogna notare che le equazioni parametriche di una superficie non sono univoche, basta infatti sostituire ad u ed a v due funzioni di altri due parametri, ad esempio s e t , per far modificare le equazioni parametriche. Concludiamo dunque che una superficie ha infinite rappresentazioni mediante equazioni parametriche.

Presentiamo come esempi di superfici immerse in uno spazio tridimensionale il **piano** e la **superficie sferica**.

Un **piano** possiede la semplice rappresentazione parametrica:

$$\begin{cases} x = au + bv + c \\ y = a'u + b'v + c' \\ z = a''u + b''v + c'' \end{cases}$$

dove $a, b, c, a', b', c', a'', b''$ e c'' sono numeri reali. Queste equazioni rappresentano un piano perché sono equazioni lineari (di primo grado) così come accade per le rette del piano, le cui equazioni sono anch'esse lineari e del tipo $ax + by + c = 0$.

Una rappresentazione della **sfera** di raggio r centrata nell'origine è la seguente, ottenuta come unione di due **semisfere**:

$$\begin{cases} x = u \\ y = v \\ z = \sqrt{r^2 - u^2 - v^2} \end{cases} \quad \wedge \quad \begin{cases} x = u \\ y = v \\ z = -\sqrt{r^2 - u^2 - v^2} \end{cases}$$

Queste sono le equazioni di una sfera, poiché risolvendo per sostituzione i due sistemi di tre equazioni ed elevando al quadrato, le soluzioni di entrambi soddisfano l'equazione $x^2 + y^2 + z^2 = r^2$, che rappresenta appunto la sfera di raggio r centrata nell'origine.

Il piano è un esempio di **spazio metrico euclideo**, ovvero uno spazio dove sia possibile definire la metrica euclidea introdotta in precedenza e basata sul teorema

di Pitagora, mentre la sfera non lo è (si ricordi il controesempio fornito prima riguardante il triangolo costruito su di una superficie sferica).

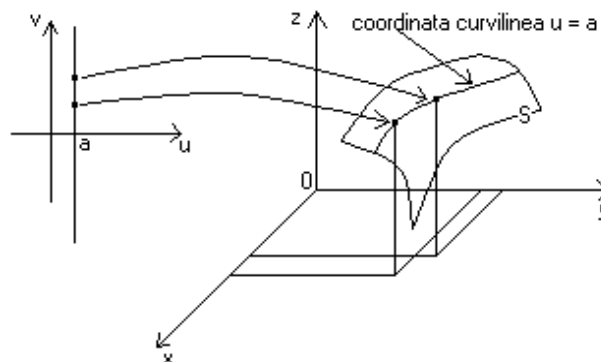
2.6 Coordinate curvilinee o generali

Dotare uno spazio euclideo di un sistema di coordinate cartesiane ortogonali è una operazione semplice ed intuitiva, tanto per lo spazio piatto a due dimensioni (un piano), che per l'intero spazio tridimensionale. Si possono immaginare anche sistemi di coordinate cartesiane ortogonali in spazi n -dimensionali come naturali estensioni di quelli a due e tre dimensioni.

Su di uno spazio curvo, invece, scegliere un sistema di coordinate può rivelarsi un'operazione più complicata. Occorre innanzitutto abbandonare il concetto di coordinate cartesiane ortogonali, ed introdurre la nozione di **coordinate curvilinee o generali**.

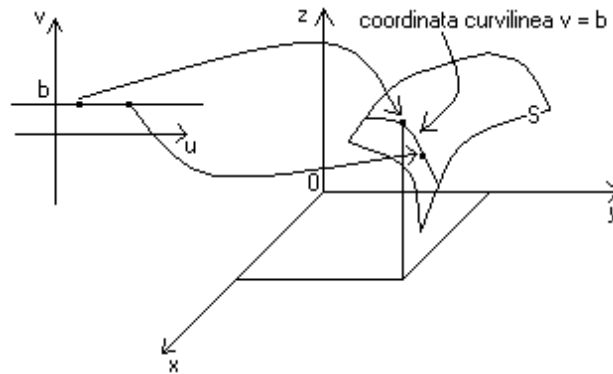
Per fare questo immaginiamo di fissare il valore del parametro u pari ad una costante reale, sia $u = a$. Se u è fisso, v può invece variare. Sul piano parametrico otterremo allora una retta parallela all'asse delle v .

Immaginiamo ora di sostituire le coppie (u, v) di questa retta con $u = a$ nell'equazione della superficie. Otterremo allora una linea sulla superficie medesima come indicato nella figura:

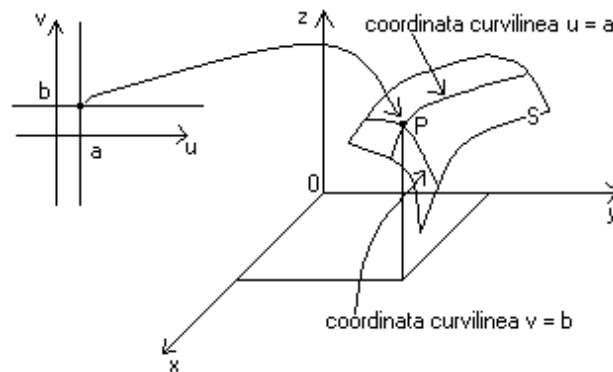


chiameremo *coordinata curvilinea* $u = a$ la linea tracciata sulla superficie in corrispondenza del parametro $u = a$.

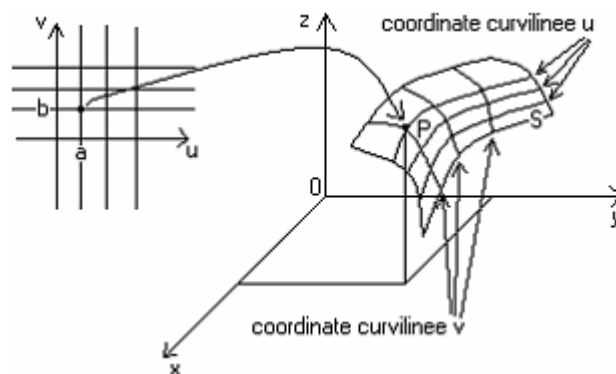
Ripetiamo la stessa operazione tenendo fisso il parametro v e facendo variare u . Se il parametro v assume il valore fisso $v = b$, otteniamo sulla superficie una linea che chiameremo *coordinata curvilinea* $v = b$ come indicato in figura:



Il punto P corrispondente ai valori dei parametri $u = a$ e $v = b$ è il punto di intersezione fra le due coordinate curvilinee corrispondenti, come indicato nel seguente grafico:



Se tracciamo più coordinate curvilinee sulla superficie otteniamo un grafico del tipo:



Possiamo dunque liberarci dal vincolo delle coordinate cartesiane ortogonali, sia per spazi di dimensione due o tre, sia per il generico spazio euclideo ad n dimensioni, facendo ricorso alle cosiddette coordinate curvilinee, o generali, ovvero tramite sistemi di equazioni leganti le coordinate originarie a quelle nuove (tali sistemi devono però poter essere invertibili, e questa richiesta è soddisfatta dalla condizione che il determinante dello Jacobiano del sistema sia non nullo).

Alcuni semplici esempi di coordinate curvilinee sono le **coordinate polari** (nel piano) e le **coordinate cilindriche** o quelle **sferiche** (nello spazio), di cui segue una presentazione, con particolare attenzione sull'effetto dell'introduzione di tali coordinate sul calcolo della distanza tra due punti supposti infinitamente vicini P e P' che, nei casi a due e tre dimensioni, in coordinate cartesiane ortogonali e con la notazione tipica della geometria differenziale può essere espressa rispettivamente come $ds^2 = dx^2 + dy^2$ e $ds^2 = dx^2 + dy^2 + dz^2$ (secondo il teorema di Pitagora e dove in questo caso gli apici indicano la potenza seconda); segue intuitivamente la generalizzazione a dimensioni superiori.

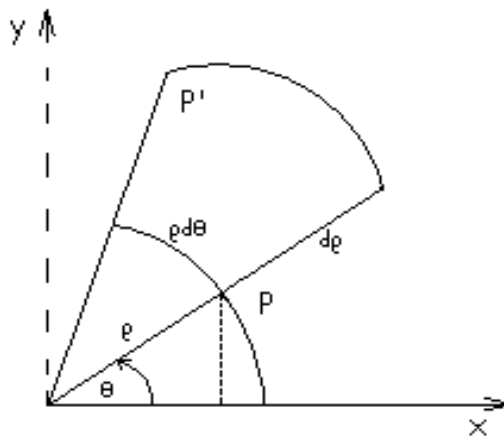
Coordinate polari

Riferiamo il piano alle coordinate polari, ovvero applichiamo la trasformazione regolata dal seguente sistema:

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \end{cases}$$

il quale è perfettamente invertibile (sotto la condizione $-\pi < \theta < \pi$) come segue:

$$\begin{cases} \rho = \sqrt{x^2 + y^2} \\ \theta = \arcsin\left(\frac{y}{\rho}\right) = \arccos\left(\frac{x}{\rho}\right) \end{cases}$$



Le linee ρ ($\theta = \text{cost.}$) sono rette uscenti da un punto, mentre le linee θ ($\rho = \text{cost.}$) sono circonferenze concentriche, aventi il centro in quel punto. Tramite questa trasformazione otteniamo che la distanza tra i due punti P e P' vale $ds^2 = d\rho^2 + \rho^2 d\theta^2$.

In generale, per riferire il piano a coordinate curvilinee generali, basterà porre le coordinate cartesiane ortogonali (x, y) eguali a funzioni regolari ed invertibili di due nuove variabili (ξ, η) , secondo un sistema invertibile del tipo

$$\begin{cases} x = x(\xi, \eta) \\ y = y(\xi, \eta) \end{cases}$$

ed in virtù di esse si avrà allora:

$$ds^2 = dx^2 + dy^2 = Ed\xi^2 + 2Fd\xi d\eta + Gd\eta^2,$$

dove E , F e G sono, in generale, funzioni di ξ e η (e più in particolare rappresentano le componenti del **tensore metrico**, che definiremo a breve). Le nuove linee coordinate sono le ξ ($\eta=\text{cost.}$) e le η ($\xi=\text{cost.}$); in generale esse saranno curvilinee ed oblique.

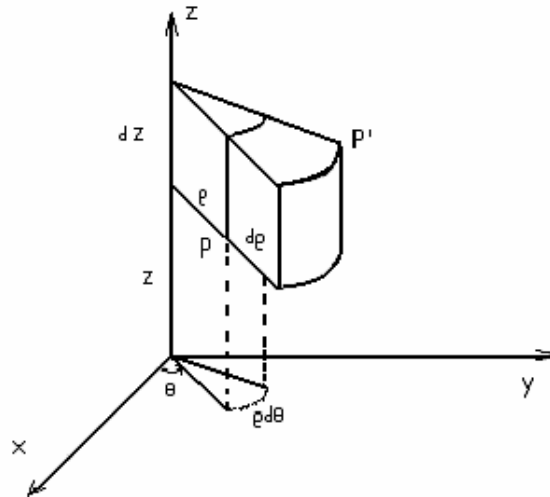
Coordinate cilindriche

Riferiamo lo spazio alle coordinate cilindriche, ovvero, in analogia a quanto fatto poc'anzi con riferimento al piano, applichiamo la trasformazione di coordinate definita dal sistema

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \\ z = z \end{cases}$$

anch'esso perfettamente invertibile:

$$\begin{cases} \rho = \sqrt{x^2 + y^2} \\ \theta = \arctan\left(\frac{y}{x}\right) = \arcsin\left(\frac{y}{\sqrt{x^2 + y^2}}\right) = \arccos\left(\frac{x}{\sqrt{x^2 + y^2}}\right) \\ z = z \end{cases}$$



Le linee ρ ($\theta = \text{cost.}$, $z = \text{cost.}$) sono rette parallele al piano xy , passanti per i punti dell'asse z e formanti un angolo θ con l'asse x ; le superfici $\rho = \text{cost.}$ sono cilindri circolari coassiali aventi z per asse; le linee θ ($\rho = \text{cost.}$, $z = \text{cost.}$) sono circonferenze in piani paralleli al piano xy e con centri sull'asse z ; le superfici $\theta = \text{cost.}$ sono piani per l'asse z ; le linee z ($\rho = \text{cost.}$, $\theta = \text{cost.}$) sono rette parallele all'asse z ; le superfici $z = \text{cost.}$ sono piani paralleli al piano xy . Sotto questa trasformazione la distanza fra i due punti P e P' vale $ds^2 = d\rho^2 + \rho^2 d\theta^2 + dz^2$.

Analogamente a quanto affermato in precedenza: in generale, per riferire lo spazio a coordinate curvilinee generali basterà porre le coordinate cartesiane ortogonali eguali a funzioni regolari ed invertibili di tre nuove variabili; lungo ciascuna delle nuove linee coordinate varierà una sola di queste variabili per volta, mentre lungo una superficie coordinata ne varieranno due, rimanendo costante la terza. E il quadrato della distanza fra due punti infinitamente vicini ds^2 risulterà, grazie a queste formule, una forma differenziale quadratica in tali nuove variabili, i cui coefficienti saranno, sempre in generale, funzioni delle variabili stesse.

2.7 Metrica di uno spazio curvo e tensore metrico fondamentale

Le considerazioni precedenti si estendono facilmente ad un $E_{(n)}$ di qualsivoglia dimensione, di cui il piano e lo spazio ordinari sono casi particolari. Per riferire uno spazio euclideo ad n dimensioni a coordinate curvilinee generali basterà dunque porre le n coordinate cartesiane ortogonali eguali a funzioni regolari ed invertibili di n nuove variabili: lungo una delle nuove linee coordinate varierà una sola di queste variabili, rimanendo costanti tutte le altre, e così lungo una ipersuperficie coordinata. Come possiamo allora definire una struttura metrica su uno spazio n -dimensionale? Ovvero, come possiamo calcolare la distanza fra due punti?

Il quadrato della distanza tra due punti infinitamente vicini, che in coordinate cartesiane ortogonali è la somma dei quadrati dei differenziali delle coordinate, diverrà, in coordinate generali, una forma differenziale quadratica in n variabili i cui coefficienti saranno, in generale, funzioni delle variabili stesse. Analizziamo ancora una volta per semplicità il caso di una superficie curva immersa in uno spazio tridimensionale, i risultati potranno essere immediatamente estesi al caso n -dimensionale.

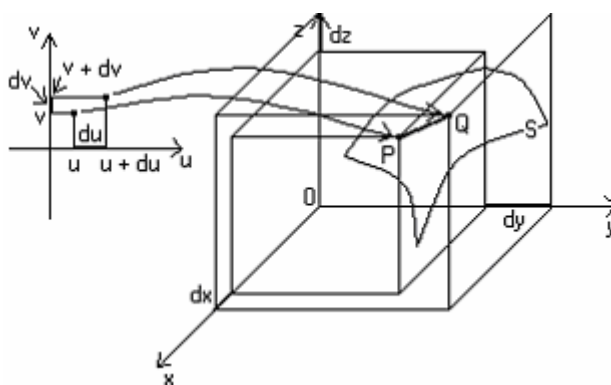
Consideriamo due punti della superficie P e Q . La loro distanza, ovviamente, deve tener conto della **curvatura** (non piattezza) della superficie per cui, calcolarla, non risulta così immediato. Nel misurare la distanza dobbiamo seguire la curvatura della superficie perché dobbiamo restare dentro lo spazio curvo che stiamo studiando.

Tracciare un segmento rettilineo PQ che si distacchi dalla superficie, cioè immerso nello spazio tridimensionale che contiene la superficie, purtroppo, ci porterebbe fuori dalla superficie e non risolverebbe il problema (anche se in questo caso la soluzione sarebbe matematicamente semplice, basterebbe impiegare il teorema di Pitagora). Per superare questa difficoltà dobbiamo fare ricorso alle idee di base del calcolo differenziale.

Supponiamo che i due punti P e Q siano molto vicini. In questo caso il segmento PQ , immerso nello spazio tridimensionale bene approssima l'arco (curvo) di superficie che unisce P con Q , e questa approssimazione è tanto migliore quanto più

vicini sono i due punti; questa semplice considerazione è l'idea chiave che ci permetterà di calcolare la distanza su ogni superficie curva. Si tratta inoltre di un'idea assai familiare, il pianeta su cui viviamo è pressoché sferico, ma entro porzioni ristrette di questa superficie essa appare a tutti gli effetti piana, per cui la distanza di due punti sulla superficie terrestre, purché sufficientemente vicini, non risente della sua curvatura; in effetti un ingegnere che progetti una casa non tiene assolutamente conto della curvatura terrestre.

Una piccola porzione di superficie curva è, quindi, praticamente piana e su tale porzione vale con grande approssimazione la geometria euclidea. Immaginiamo allora di considerare il punto P corrispondente ai valori dei parametri u e v ed il punto Q ai valori $u+du$ e $v+dv$, ove con du e dv nel calcolo differenziale indichiamo un incremento infinitesimo sulle variabili u e v rispettivamente. La lunghezza del segmento PQ (preso nello spazio tridimensionale che contiene la superficie) allora rappresenta la lunghezza del segmento rettilineo tracciato fra i due punti sulla superficie; le variazioni delle coordinate (x, y, z) vengono analogamente indicate con dx, dy e dz . Tutto ciò è rappresentato nel grafico seguente:



Applicando il teorema di Pitagora otteniamo che la distanza ds^2 risulta essere

$$ds^2 = dx^2 + dy^2 + dz^2.$$

La lunghezza al quadrato così trovata, come sopra precisato, non corrisponde alla lunghezza dell'arco di superficie compreso fra i punti P e Q , ne costituisce solo una

approssimazione, tanto migliore quanto più essi sono vicini. Questa lunghezza, però, corrisponderebbe esattamente alla distanza fra i punti P e Q (considerati come punti dello spazio tridimensionale) anche se gli incrementi dx, dy e dz non fossero infinitesimi, ovvero varrebbe anche se P e Q fossero molto distanti tra di loro. L'espressione di questa lunghezza dipende, come indicato dalla formula, solo dalle coordinate cartesiane ortogonali dello spazio euclideo in cui la superficie è immersa. Non vi è nessuna informazione circa la superficie e le sue coordinate curvilinee; occorre dunque esprimere ds^2 in funzione dei parametri u e v . Solo così è possibile costruire una metrica sulla superficie che tenga conto della sua curvatura (ovvero delle sue proprietà intrinseche). Così facendo, nella trattazione matematica della superficie rimangono solo le sue coordinate curvilinee, e null'altro; questo fatto è fondamentale e costituisce l'anima vera e propria del calcolo tensoriale.

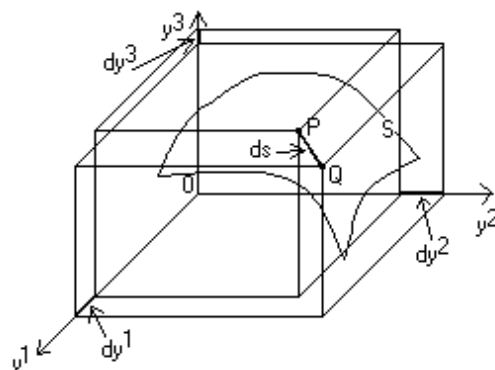
Mostriamo dunque come, con l'ausilio del calcolo differenziale, sia possibile introdurre le coordinate curvilinee in luogo di quelle cartesiane ortogonali, e definiamo rigorosamente una metrica sulla superficie curva che tenga conto delle sue proprietà intrinseche.

Consideriamo una superficie S immersa in un $E_{(3)}$ le cui equazioni parametriche sono, come è noto,

$$\begin{cases} y^1 = y^1(x^1, x^2) \\ y^2 = y^2(x^1, x^2) \\ y^3 = y^3(x^1, x^2) \end{cases}$$

ove y^h , $h=1,2,3$ indicano le coordinate nel sistema di riferimento cartesiano ortogonale, mentre x^k , $k=1,2$ sono i parametri (gli indici ed i numeri ad apice sulle coordinate e sui parametri non indicano elevamento a potenza), e si è ripresa la notazione tipica del calcolo tensoriale per compattezza e comodità, varrà sempre la convenzione sulla somma degli indici ripetuti.

Si consideri ora un punto P sulla superficie S di coordinate $P(y^1, y^2, y^3)$ ed un punto Q , anch'esso sulla superficie, a P infinitamente vicino e di coordinate $P(y^1 + dy^1, y^2 + dy^2, y^3 + dy^3)$ il che si traduce graficamente come segue:



La lunghezza del segmento infinitesimo PQ è la distanza fra i due punti, e viene solitamente chiamata **elemento di linea**, ed indicata, come già abbiamo usato fare, con ds . Tale elemento di linea è uno scalare **invariante** perché, pur cambiando il sistema di riferimento, esso non si modifica. Il segmento PQ tende inoltre a giacere completamente sulla superficie quanto più i due punti sono tra loro vicini; l'elemento ds è quindi considerabile come appartenente alla superficie, ovvero come formante un suo **arco infinitesimo**; applicando infine il teorema di Pitagora risulta

$$ds^2 = (dy^1)^2 + (dy^2)^2 + (dy^3)^2.$$

Esprimiamo ora i differenziali delle variabili y^h , $h=1,2,3$, in funzione dei differenziali delle variabili x^k , $k=1,2$, tramite le note formule del calcolo differenziale; avremo quindi che

$$dy^h = \frac{\partial y^h}{\partial x^k} dx^k, \quad h=1,2,3, \quad k=1,2$$

(è sottintesa la sommatoria sugli indici ripetuti). Questo ci porta al seguente sistema:

$$\begin{cases} dy^1 = \frac{\partial y^1}{\partial x^1} dx^1 + \frac{\partial y^1}{\partial x^2} dx^2 \\ dy^2 = \frac{\partial y^2}{\partial x^1} dx^1 + \frac{\partial y^2}{\partial x^2} dx^2 \\ dy^3 = \frac{\partial y^3}{\partial x^1} dx^1 + \frac{\partial y^3}{\partial x^2} dx^2 \end{cases}$$

Ora siamo in grado di ricavare il quadrato dell'elemento di linea ds^2 sostituendo le relazioni trovate nella formula $ds^2 = (dy^1)^2 + (dy^2)^2 + (dy^3)^2$ (che fornisce, come già detto, la distanza fra due punti infinitamente vicini nello spazio tridimensionale euclideo) e svolgendo opportune semplificazioni; si ottiene la seguente formula:

$$ds^2 = \left[\left(\frac{\partial y^1}{\partial x^1} \right)^2 + \left(\frac{\partial y^2}{\partial x^1} \right)^2 + \left(\frac{\partial y^3}{\partial x^1} \right)^2 \right] (dx^1)^2 + 2 \left(\frac{\partial y^1}{\partial x^1} \frac{\partial y^1}{\partial x^2} + \frac{\partial y^2}{\partial x^1} \frac{\partial y^2}{\partial x^2} + \frac{\partial y^3}{\partial x^1} \frac{\partial y^3}{\partial x^2} \right) dx^1 dx^2 + \left[\left(\frac{\partial y^1}{\partial x^2} \right)^2 + \left(\frac{\partial y^2}{\partial x^2} \right)^2 + \left(\frac{\partial y^3}{\partial x^2} \right)^2 \right] (dx^2)^2$$

Si tratta indubbiamente di una espressione ingombrante, che può però essere resa più compatta introducendo la matrice

$$\mathcal{G} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$

dove

$$\begin{cases} g_{11} = \left(\frac{\partial y^1}{\partial x^1} \right)^2 + \left(\frac{\partial y^2}{\partial x^1} \right)^2 + \left(\frac{\partial y^3}{\partial x^1} \right)^2 \\ g_{12} = g_{21} = \frac{\partial y^1}{\partial x^1} \frac{\partial y^1}{\partial x^2} + \frac{\partial y^2}{\partial x^1} \frac{\partial y^2}{\partial x^2} + \frac{\partial y^3}{\partial x^1} \frac{\partial y^3}{\partial x^2} \\ g_{22} = \left(\frac{\partial y^1}{\partial x^2} \right)^2 + \left(\frac{\partial y^2}{\partial x^2} \right)^2 + \left(\frac{\partial y^3}{\partial x^2} \right)^2 \end{cases}$$

ovvero, in modo compatto, l'elemento (i, j) -esimo di tale matrice è

$$g_{ij} = \frac{\partial y^h}{\partial x^i} \frac{\partial y^h}{\partial x^j} \quad h = 1, 2, 3, \quad i, j = 1, 2.$$

Con questa sostituzione l'espressione di ds^2 diventa di più agevole lettura:

$$ds^2 = g_{11} (dx^1)^2 + 2g_{12} dx^1 dx^2 + g_{22} (dx^2)^2$$

o, equivalentemente, sfruttando la notazione tensoriale e la convenzione sulla somma degli indici ripetuti (è immediato vedere come gli indici si saturino dando luogo ad uno scalare, la distanza al quadrato appunto):

$$ds^2 = g_{ij} dx^i dx^j \quad i, j = 1, 2.$$

La distanza così definita prende il nome di **metrica di Riemann**, e le grandezze g_{ij} introdotte formano il cosiddetto **tensore metrico fondamentale**.

Tale tensore contiene in sé tutte le informazioni relative alle proprietà metriche della superficie curva, ed è espresso in funzione delle sole coordinate curvilinee, permettendo così di lavorare sullo spazio curvo senza avere la necessità di considerarlo immerso in uno spazio euclideo di dimensione maggiore; dalla sola conoscenza del tensore metrico è possibile risalire alle **geodetiche** (linee di minima lunghezza) dello spazio curvo oggetto di studio, così come alla **curvatura**. Inoltre il tensore metrico è di ordine due, simmetrico e definito positivo (condizione questa senza la quale l'elemento ds^2 potrebbe essere negativo, ciò discende dalla teoria delle forme quadratiche). Il discorso è generalizzabile ad n dimensioni, ovvero facendo riferimento a spazi curvi n -dimensionali, non più necessariamente solo superfici, bensì ad una qualsivoglia varietà differenziabile, la quale, dotata di una metrica come quella appena descritta, viene definita **varietà riemanniana**.

Definiamo in modo più rigoroso una varietà riemanniana come una restrizione di uno spazio euclideo. Siano y_α , ($\alpha = 1, 2, \dots, n$) coordinate cartesiane ortogonali in un $E_{(n)}$, e poniamo

$$y_\alpha = y_\alpha(x^i), \quad \begin{pmatrix} i = 1, 2, \dots, m \\ m < n \end{pmatrix}$$

dove le y_α sono funzioni regolari dei loro argomenti. Il punto $P(y_\alpha)$ (o se si vuole il vettore $P-O$ con O punto prefissato) diverrà funzione delle nuove variabili:

$$P = P(x^i),$$

e al variare di queste ultime descriverà una varietà $V_{(m)}$ che diremo *immersa* nell' $E_{(n)}$ euclideo per assunzione. Le linee x^i , ottenute facendo variare una sola di queste variabili e dando a tutte le altre valori costanti, costituiscono un sistema di coordinate curvilinee sopra la varietà $V_{(m)}$.

Il quadrato della distanza tra due punti infinitamente vicini di $V_{(m)}$, $P(x^i)$ e $P(x^i + dx^i)$, che, nello spazio euclideo ambiente, in coordinate cartesiane ortogonali, ha la forma pitagorica in n variabili

$$ds^2 = dP \times dP = \sum_{\alpha=1}^n dy_\alpha^2,$$

diverrà, in forza delle relazioni fra le y_α e le x^i , una forma differenziale quadratica nelle m variabili x^i :

$$(8) \quad ds^2 = g_{ik} dx^i dx^k$$

dove

$$g_{ik} = g_{ki} = \frac{\partial P}{\partial x^i} \times \frac{\partial P}{\partial x^k} = \frac{\partial y_\alpha}{\partial x^i} \frac{\partial y_\alpha}{\partial x^k}$$

altro non sono che i coefficienti del tensore metrico (è sempre sottintesa la somma rispetto agli indici ripetuti); tale forma quadratica è quindi definita positiva, come già affermato. L'espressione (8) è detta *metrica* di $V_{(m)}$. In generale non è possibile scegliere m variabili tali che essa assuma la forma pitagorica. Quando questo è possibile la $V_{(m)}$ si dice *euclidea* ed abbiamo così il caso in cui le componenti g_{ik} sono tutte costanti e le coordinate della varietà sono cartesiane ortogonali, ma questo è solo un caso particolare; in generale il tensore metrico non è costante in tutte le sue componenti e dipende, come già detto, dalle coordinate curvilinee definite sulla varietà. Si parlerà allora in generale di **varietà riemanniana**; se, in particolare, $m = n - 1$, la $V_{(m)}$ si dice *ipersuperficie* di $E_{(n)}$, se $m = 2$ semplicemente *superficie*.

Le superfici dello spazio ordinario sono dunque il più semplice esempio di varietà riemanniane a due dimensioni e, per quanto detto, le varietà euclidee sono casi particolari di varietà riemanniane. Così, nel caso di due dimensioni, il piano e il cilindro sono varietà euclidee, mentre non lo è la sfera, la cui metrica non assume la forma pitagorica (ovvero senza elementi di “interazione” fra le coordinate).

Si presenta però la questione della liceità o meno di imporre un qualsivoglia sistema di coordinate sulla varietà, tipicamente dimensionalmente inferiore allo spazio euclideo in cui si trova immersa, a fronte della necessità dell'identificabilità delle condizioni sperimentali e dei risultati. Nei sistemi di coordinate esemplificativi presentati, polari e cilindriche, tale questione non si poneva, in forza della perfetta invertibilità dei sistemi definitivi (il numero di dimensioni dello spazio di partenza e di quello di arrivo coincidevano e le funzioni erano regolari ed invertibili); più in generale però non ci si troverà in questa situazione, bensì si avranno equazioni definitive della varietà in numero non sufficiente per garantire l'invertibilità globale del sistema. Infatti risulta evidente come, a fronte di sole $k < n$ equazioni definitive del sistema di coordinate generali su di una varietà $V_{(k)}$, non sia possibile risalire univocamente allo spazio $E_{(n)}$ in cui la varietà è immersa, sarà necessario fissare arbitrariamente $n - k$ coordinate ulteriori per poterlo fare, e sempre sotto opportune condizioni di regolarità; tale problematica può essere superata per i nostri intenti facendo ricorso ad un risultato denominato *Injective Mapping Theorem*, il

quale in sostanza afferma che a fronte di condizioni di regolarità ed iniettività delle derivate prime delle relazioni funzionali che ci interessano, è garantita l'invertibilità *locale*, fatto salvo che nei punti critici (che non sono oggetto di studio nel presente lavoro). Tali condizioni possono tradursi nel richiedere che il determinante della matrice Jacobiana associata al sistema di trasformazione di coordinate sia di rango massimo, ovvero che esista un minore invertibile di ordine k . Grazie a questo risultato, e supponendo di lavorare sempre nelle condizioni di regolarità richieste, possiamo considerare valida la definizione di un sistema di coordinate curvilinee a cui ci appoggeremo per ricavare le informazioni necessarie sulla varietà oggetto di studio.

2.8 Interpretazione di un modello non lineare come una varietà riemanniana

In conclusione di capitolo vediamo quindi come interpretare un modello probabilistico generico in chiave geometrica, ricordando l'impostazione utilizzata per il modello di regressione lineare nei parametri, che alla luce di quanto visto si presenta dunque come uno spazio euclideo, in contrapposizione ai modelli non lineari i quali, come vedremo, costituiranno delle varietà riemanniane.

Sia dato il modello a fattori controllabili (e supposti fissi) non lineare nei parametri

$$y_t = f(\mathbf{X}_t, \boldsymbol{\theta}) + \varepsilon_t$$

in cui (riprendendo dal paragrafo 1.1) $\boldsymbol{\theta} = (\vartheta_1, \vartheta_2, \dots, \vartheta_p)'$ è l'insieme degli ignoti parametri oggetto di stima, \mathbf{X}_t la matrice contenente le condizioni sperimentali ed $\varepsilon_t, t = 1, \dots, n$, è la componente erratica (supposta additiva) di natura stocastica. Se gli errori soddisfano le condizioni

$$\begin{aligned} E[\varepsilon_t] &= 0, & t &= 1, \dots, n \\ E[\varepsilon_t, \varepsilon_s] &= \sigma^2 \delta_{ts}, & t, s &= 1, \dots, n \end{aligned}$$

dove δ_{ts} è il delta di Kronecker (che assume valore 1 quando $t = s$, 0 altrimenti), allora è possibile, come già detto, ottenere le stime dei parametri secondo il metodo dei LS.

Scrivendo in forma estesa le medie $\eta(\boldsymbol{\theta})$ del modello, supposto di avere ottenuto, a fronte di una sperimentazione, e per un dato insieme di condizioni sperimentali, una collezione di risposte $y_t, t = 1, \dots, n$, abbiamo:

$$\begin{cases} \eta_1(\boldsymbol{\theta}) = f(\mathbf{X}_1, \boldsymbol{\theta}) \\ \eta_2(\boldsymbol{\theta}) = f(\mathbf{X}_1, \boldsymbol{\theta}) \\ \dots \\ \eta_n(\boldsymbol{\theta}) = f(\mathbf{X}_1, \boldsymbol{\theta}) \end{cases}$$

Questo sistema può essere visto come l'insieme di relazioni definitorie di una restrizione k -dimensionale (ovvero di dimensione pari al numero di parametri) di uno spazio euclideo n -dimensionale (ovvero di dimensione pari al numero delle osservazioni), ovvero, per quanto detto, una varietà riemanniana $V_{(k)}$ immersa in un $E_{(n)}$.

Per verificare se la varietà si riduce ad una varietà lineare (euclidea) è sufficiente calcolare il tensore metrico, e verificare se la forma quadratica ad esso associato (la metrica della varietà) assuma la forma pitagorica; affinché sia vero è sufficiente verificare che questo tensore abbia componenti costanti.

Prendiamo in esame due semplici modelli, uno lineare e uno non lineare, con $n=3$ e $k=2$, e le assunzioni sugli errori sono le medesime già specificate in entrambi i casi (si conserva la convenzione di sommare gli indici ripetuti).

Modello lineare

Scriviamo il modello in forma compatta come $y_t = x_{tj}\theta_j + \varepsilon_t$, e con $t=1,2,3; j=1,2$.

Avremo quindi

$$\begin{cases} \eta_1(\boldsymbol{\theta}) = x_{11}\theta_1 + x_{12}\theta_2 \\ \eta_2(\boldsymbol{\theta}) = x_{21}\theta_1 + x_{22}\theta_2 \\ \eta_3(\boldsymbol{\theta}) = x_{31}\theta_1 + x_{32}\theta_2 \end{cases}$$

Queste relazioni definiscono, come già visto, una varietà lineare di dimensione 2 immersa in uno spazio euclideo di dimensione 3.

Calcoliamo il tensore metrico, osservando che può essere ricavato semplicemente dalla matrice jacobiana come segue, essendo

$$J = \begin{bmatrix} \frac{\partial \eta_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_1(\boldsymbol{\theta})}{\partial \theta_2} \\ \frac{\partial \eta_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_2(\boldsymbol{\theta})}{\partial \theta_2} \\ \frac{\partial \eta_3(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_3(\boldsymbol{\theta})}{\partial \theta_2} \end{bmatrix}$$

la matrice jacobiana, allora

$$\begin{aligned} J'J &= \begin{bmatrix} \frac{\partial \eta_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_3(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial \eta_1(\boldsymbol{\theta})}{\partial \theta_2} & \frac{\partial \eta_2(\boldsymbol{\theta})}{\partial \theta_2} & \frac{\partial \eta_3(\boldsymbol{\theta})}{\partial \theta_2} \end{bmatrix} \begin{bmatrix} \frac{\partial \eta_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_1(\boldsymbol{\theta})}{\partial \theta_2} \\ \frac{\partial \eta_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_2(\boldsymbol{\theta})}{\partial \theta_2} \\ \frac{\partial \eta_3(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial \eta_3(\boldsymbol{\theta})}{\partial \theta_2} \end{bmatrix} = \\ &= \begin{bmatrix} \sum_t \left(\frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \theta_1} \right)^2 & \frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \theta_1} \frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \theta_2} \\ \frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \theta_1} \frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \theta_2} & \sum_t \left(\frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \theta_2} \right)^2 \end{bmatrix} = \frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \theta_j} = [g_{ij}] = \mathcal{G} \end{aligned}$$

Calcolando lo jacobiano del sistema otteniamo

$$J = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix}$$

da cui

$$J'J = \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix} = \begin{bmatrix} x_{11}^2 + x_{21}^2 + x_{31}^2 & x_{11}x_{12} + x_{21}x_{22} + x_{31}x_{32} \\ x_{11}x_{12} + x_{21}x_{22} + x_{31}x_{32} & x_{12}^2 + x_{22}^2 + x_{32}^2 \end{bmatrix} = \mathcal{G}$$

matrice la quale, non dipendendo dai parametri, ed essendo le x_{ij} costanti assegnate (sono le condizioni sperimentali) risulta costante.

Modello non lineare

Si consideri ora il seguente semplice modello non lineare nei parametri, che trascriviamo già in termini di medie:

$$\begin{cases} \eta_1(\boldsymbol{\theta}) = x_{11}^{\theta_1} + x_{12}^{\theta_2} \\ \eta_2(\boldsymbol{\theta}) = x_{21}^{\theta_1} + x_{22}^{\theta_2} \\ \eta_3(\boldsymbol{\theta}) = x_{31}^{\theta_1} + x_{32}^{\theta_2} \end{cases}$$

Analogamente a quanto fatto prima calcoliamo la matrice jacobiana e quindi il tensore metrico:

$$J = \begin{bmatrix} \ln(x_{11})x_{11}^{\theta_1} & \ln(x_{12})x_{12}^{\theta_2} \\ \ln(x_{21})x_{21}^{\theta_1} & \ln(x_{22})x_{22}^{\theta_2} \\ \ln(x_{31})x_{31}^{\theta_1} & \ln(x_{32})x_{32}^{\theta_2} \end{bmatrix}$$

$$J'J = \begin{bmatrix} \ln(x_{11})x_{11}^{\theta_1} & \ln(x_{21})x_{21}^{\theta_1} & \ln(x_{31})x_{31}^{\theta_1} \\ \ln(x_{12})x_{12}^{\theta_2} & \ln(x_{22})x_{22}^{\theta_2} & \ln(x_{32})x_{32}^{\theta_2} \end{bmatrix} \begin{bmatrix} \ln(x_{11})x_{11}^{\theta_1} & \ln(x_{12})x_{12}^{\theta_2} \\ \ln(x_{21})x_{21}^{\theta_1} & \ln(x_{22})x_{22}^{\theta_2} \\ \ln(x_{31})x_{31}^{\theta_1} & \ln(x_{32})x_{32}^{\theta_2} \end{bmatrix} = \mathcal{G}$$

con

$$\begin{aligned}
g_{11} &= \ln^2(x_{11})x_{11}^{2\theta_1} + \ln^2(x_{21})x_{21}^{2\theta_1} + \ln^2(x_{31})x_{31}^{2\theta_1} \\
g_{12} = g_{21} &= \ln(x_{11})x_{11}^{\theta_1} \ln(x_{12})x_{12}^{\theta_2} + \ln(x_{21})x_{21}^{\theta_1} \ln(x_{22})x_{22}^{\theta_2} + \ln(x_{31})x_{31}^{\theta_1} \ln(x_{32})x_{32}^{\theta_2} \\
g_{22} &= \ln^2(x_{12})x_{12}^{2\theta_2} + \ln^2(x_{22})x_{22}^{2\theta_2} + \ln^2(x_{32})x_{32}^{2\theta_2}
\end{aligned}$$

Si può chiaramente vedere come le componenti del tensore metrico dipendano dai parametri θ_1, θ_2 , e non risultino dunque costanti; lo jacobiano risulta inoltre di rango massimo, per le condizioni di regolarità in cui ci siamo posti.

Alla luce di questa verifica possiamo affermare quindi che le medie del modello non lineare considerato (ed è più in generale vero per qualsivoglia modello non lineare nei parametri) costituiscano a tutti gli effetti una varietà riemanniana.

CAPITOLO III

Il concetto di curvatura

Il termine **curvatura** indica una serie di concetti geometrici legati fra di loro, che intuitivamente si riferiscono alla misura di quanto un determinato oggetto si discosti dall'essere piatto; in tal senso, la curvatura viene definita in modi differenti a seconda dell'ente geometrico cui è riferita. La nozione di curvatura è alla base della Geometria Differenziale, ed ha notevoli applicazioni in fisica teorica (in particolare nella relatività generale) ma, come vedremo, si dimostra essere un utile strumento anche in campo statistico per quanto riguarda l'analisi dei modelli non lineari. Possiamo già quindi anticipare che saremo interessati alla curvatura del luogo delle soluzioni di un modello non lineare nell'intorno del punto di stima parametrica, ed in particolare a quanto questo dipartirsi dalla situazione di piattezza (ovvero euclidea) influenzi le proprietà degli stimatori.

Curvatura estrinseca ed intrinseca

Si distinguono due tipologie essenziali di curvatura:

- Curvatura **estrinseca**: è la curvatura posseduta dall'oggetto in relazione allo spazio piatto di dimensione superiore in cui è immerso, e determinabile solo confrontando gli elementi dell'oggetto in relazione a quelli dello spazio che lo contiene.
- Curvatura **intrinseca**: è la curvatura determinabile utilizzando solo operazioni eseguite su elementi dell'oggetto medesimo; si tratta di una di quelle proprietà intrinseche delle varietà cui si era fatto riferimento nel secondo capitolo.

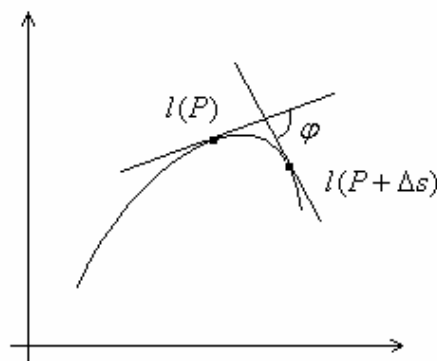
Un esempio di curvatura estrinseca è quella di una superficie cilindrica nello spazio tridimensionale: le linee tracciate sul cilindro sono curve se confrontate con le rette

dello spazio in cui è immerso, ma la geometria intrinseca di tale superficie è invece piatta, in quanto su di essa valgono tutti gli assiomi euclidei (la metrica assume forma pitagorica).

Una sfera è invece dotata di una curvatura intrinseca, determinabile rimanendo all'interno della superficie stessa. Si ricordi l'esempio del paragrafo 2.5 relativo ad un punto in movimento sulla superficie terrestre (accettabilmente approssimabile da una sfera), un analogo percorso eseguito su un piano non riporterà mai al punto di partenza, infatti viene violato l'assioma secondo il quale la somma degli angoli interni di un triangolo è sempre pari a 180° .

3.1 Curvatura di una linea nel piano

Consideriamo una curva l in \mathbb{R}^2 , ovvero una varietà 1-dimensionale immersa in uno spazio euclideo a due dimensioni; sottolineiamo che parleremo d'ora in poi di curvatura in un punto (di una curva adesso, di una superficie o di uno spazio in seguito). In questo caso la curvatura è intuitivamente la misura di quanto esso devia rispetto alla tangente alla curva nel punto dato, sarà quindi una proprietà costante **localmente**, il che significa che, salvo casi particolari, la curvatura varierà di punto in punto.



La curvatura in un punto P si calcola come

$$\kappa = \lim_{\Delta s \rightarrow 0} \frac{\Delta \varphi}{\Delta s}$$

ovvero come il limite del rapporto fra la variazione $\Delta \varphi$ dell'angolo formato dalle tangenti alla curva nei due punti P e $P + \Delta s$ al tendere di Δs a zero, ovvero all'avvicinarsi al limite alla coincidenza dei due punti. Risulta evidente come, in base a questa definizione, la retta abbia curvatura nulla.

La circonferenza offre il modello più semplice di misura della curvatura: circonferenze con raggio maggiore hanno curvatura minore, e viceversa, si dimostra infatti che la curvatura di una circonferenza è costante in ogni suo punto e pari all'inverso del suo raggio r . La retta può localmente identificarsi con la circonferenza di raggio infinito, da cui $\kappa = 0$. È il motivo per cui sul nostro pianeta, nonostante sia dotato di una curvatura, noi non la percepiamo, vivendo in una regione sufficientemente piccola rispetto alla superficie totale.

Tale considerazione può estendersi a oggetti più complessi e di dimensione maggiore, come vedremo in seguito.

Il cerchio osculatore

Per misurare la curvatura κ dovremmo quindi procedere con un passaggio al limite, che è un'operazione non sempre immediata da eseguire. Fortunatamente esiste una formula esplicita per determinarla; a tale scopo consideriamo due punti A e B che individuano un tratto di lunghezza Δt a cavallo del punto P . Se costruiamo la circonferenza passante per questi tre punti, essa approssimerà la linea sempre meglio man mano che Δt tenderà a zero, dunque se conosciamo il raggio r di questa circonferenza possiamo utilizzare la formula $\kappa = 1/r$. Trovare questa circonferenza è molto semplice, in quanto si tratta del **cerchio osculatore** (il nome è dovuto a Leibniz, il quale la definiva *circulum osculans*, ovvero circonferenza che

bacia la linea nel punto di interesse). La formula per trovare il raggio del cerchio osculatore prevede il calcolo della derivata prima e seconda della curva nel punto considerato, sostituendo nell'espressione per ottenere la curvatura otteniamo quindi le seguenti espressioni, divise a seconda che la curva sia data in forma parametrica, implicita o esplicita:

- Curva in forma parametrica $\alpha(t) = (x(t), y(t))$

$$\kappa = \left\| \frac{\frac{\partial x}{\partial t} \frac{\partial^2 y}{\partial t^2} - \frac{\partial y}{\partial t} \frac{\partial^2 x}{\partial t^2}}{\left(\left(\frac{\partial x}{\partial t} \right)^2 + \left(\frac{\partial y}{\partial t} \right)^2 \right)^{3/2}} \right\|$$

- Curva in forma implicita $f(x, y) = 0$

$$\kappa = \left\| \nabla \left(\frac{\nabla f}{\|\nabla f\|} \right) \right\|$$

dove ∇ è l'operatore **nabla** (che è un operatore differenziale tensoriale), che permette di scrivere in modo più compatto diverse relazioni della Geometria Differenziale, in particolare applicandolo ad un campo scalare (come nel linguaggio tecnicamente più corretto è da definirsi una funzione) se ne ottiene il gradiente (ovvero il vettore che ha per componenti le derivate parziali prime rispetto agli assi coordinati); inoltre $\nabla f / \|\nabla f\|$ è un vettore (più precisamente un campo vettoriale), ed applicando il nabla ad esso otteniamo la **divergenza** della direzione del gradiente.

- Curva in forma esplicita $y = f(x)$

$$\kappa = \frac{\frac{\partial^2 y}{\partial x^2}}{\left(1 + \left(\frac{\partial y}{\partial x}\right)^2\right)^{3/2}}$$

Si nota come queste quantità possano risultare negative, in contrasto con il fatto che il raggio di una circonferenza sia una grandezza positiva, ma se non ci preoccupiamo di questo aspetto, il segno di r indica semplicemente il verso della concavità della curva rispetto al nostro sistema di riferimento.

3.2 Curvatura di una linea nello spazio

Il comportamento di una curva $\alpha(t) = (x(t), y(t), z(t))$ nello spazio \mathbb{R}^3 si può descrivere appoggiandosi alla **terna di Frenet**, un sistema di riferimento formato da tre vettori ortonormali, $\mathbf{e}_1(t), \mathbf{e}_2(t), \mathbf{e}_3(t)$, dipendenti da t . Per definire tale sistema di riferimento è necessario che la curva sia regolare, condizione in cui ci poniamo per gli obiettivi di questa trattazione, non incontrando dunque problemi di sorta in questo senso. Nel piano il primo versore di Frenet è la tangente alla curva in t , mentre il secondo è il versore **normale** ad $\alpha(t)$ nella direzione in cui curva. Si definiscono tramite questi vettori ortonormali le cosiddette **curvature generalizzate**:

$$\chi_i = \frac{\langle \mathbf{e}'_i(t), \mathbf{e}_{i+1}(t) \rangle}{\|\alpha'(t)\|}$$

dove l'indice i varia da 1 al numero di dimensioni dello spazio ambiente meno uno, nel caso del piano i assume solo valore 1, nel caso dello spazio $i = 1, 2$, e così via al

crescere delle dimensioni, e con $\alpha'(t)$ indichiamo il vettore delle derivate prime di $\alpha(t)$.

Nello spazio tridimensionale i versori di Frenet e le curvatures generalizzate hanno dei nomi specifici:

Versore tangente

Il primo vettore di Frenet è il versore tangente, definito quindi come: $\mathbf{e}_1(t) = \frac{\alpha'(t)}{\|\alpha'(t)\|}$

dove $\|\alpha'(t)\|$ è la norma del vettore tangente, ovvero la radice quadrata della somma dei quadrati delle componenti di detto vettore.

Versore normale

Il versore normale misura quanto la curva differisca da una linea retta, ed è il secondo vettore di Frenet, definito come

$$\mathbf{e}_2(t) = \frac{\bar{\mathbf{e}}_2(t)}{\|\bar{\mathbf{e}}_2(t)\|}, \text{ con } \bar{\mathbf{e}}_2(t) = \alpha''(t) - \langle \alpha''(t), \mathbf{e}_1(t) \rangle \mathbf{e}_1(t)$$

dove $\alpha''(t)$ è il vettore delle derivate seconde, e il vettore $\bar{\mathbf{e}}_2(t)$ viene costruito a partire da $\mathbf{e}_1(t)$ tramite la procedura di ortogonalizzazione di Graham-Schmidt.

I versori tangente e normale formano la base per un piano, chiamato **piano osculatore** della curva nel punto t .

Prima curvatura generalizzata

La prima curvatura generalizzata $\chi_1(t)$ è chiamata semplicemente curvatura di α in t ed è data da:

$$\kappa(t) = \chi_1 = \frac{\langle \mathbf{e}'_1(t), \mathbf{e}_2(t) \rangle}{\|\alpha'(t)\|}$$

e come già detto il reciproco di tale quantità è il raggio di curvatura nel punto t .

Versore binormale

Il versore **binormale** è il terzo vettore di Frenet, ed insieme ai primi due definisce la terna denominata anche **triedro fondamentale**. Si ricava semplicemente facendo il prodotto esterno (o prodotto vettoriale) fra i primi due versori, ed è pertanto ortogonale al piano osculatore:

$$\mathbf{e}_3(t) = \mathbf{e}_1(t) \wedge \mathbf{e}_2(t)$$

Ricordiamo brevemente che il prodotto esterno è definito come segue:

Definizione 3.3.1

Siano $\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ e $\mathbf{w} = w_1\mathbf{i} + w_2\mathbf{j} + w_3\mathbf{k}$ due vettori di \mathbb{R}^3 , espressi per componenti rispetto alla base ortonormale canonica $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$. Si dice prodotto esterno di \mathbf{v} per \mathbf{w} il vettore di \mathbb{R}^3 , indicato con $\mathbf{v} \wedge \mathbf{w}$, così definito:

$$\mathbf{v} \wedge \mathbf{w} := (v_2w_3 - v_3w_2)\mathbf{i} + (v_3w_1 - v_1w_3)\mathbf{j} + (v_1w_2 - v_2w_1)\mathbf{k}$$

Si può ricordare facilmente tale espressione osservando che il secondo membro coincide formalmente con il determinante della matrice seguente, ottenuto per sviluppo secondo la prima riga:

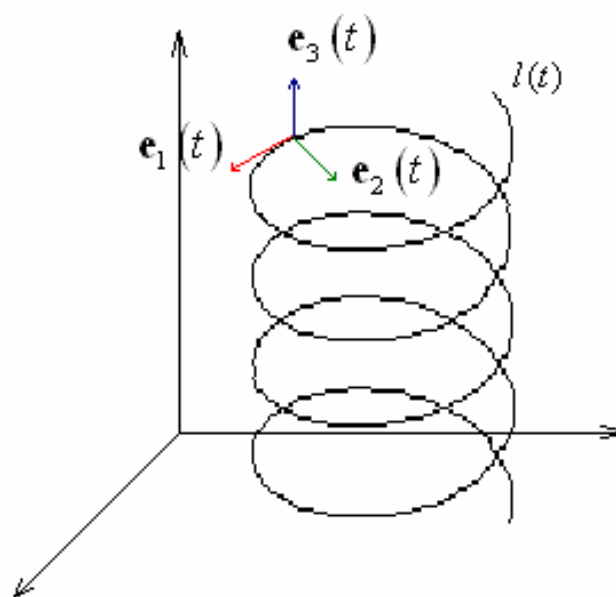
$$\begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}$$

Si fa notare come, in particolare, $\mathbf{i} \wedge \mathbf{j} = \mathbf{k}$, $\mathbf{j} \wedge \mathbf{k} = \mathbf{i}$, $\mathbf{k} \wedge \mathbf{i} = \mathbf{j}$.

Seconda curvatura generalizzata

La seconda curvatura generalizzata $\chi_2(t)$ è chiamata **torsione** e misura quanto la curva esce dal piano osculatore. Dunque una linea ha torsione nulla se e solo se è piana (ovvero esiste un piano che la contenga). Definiamo la torsione come segue:

$$\tau(t) = \chi_2 = \frac{\langle \mathbf{e}'_2(t), \mathbf{e}_3(t) \rangle}{\|\alpha'(t)\|}$$



Più in generale possiamo definire i vettori di Frenet e le curvatures generalizzate per una qualsivoglia dimensione tramite le formule di Frenet-Serret, le quali sono delle equazioni differenziali ordinarie del primo ordine, la cui soluzione è il sistema di Frenet che descrive la curva; i coefficienti dell'equazione sono dati dalle curvatures generalizzate χ_i :

$$\mathbf{2\ dimensions:} \quad \begin{bmatrix} \mathbf{e}'_1(t) \\ \mathbf{e}'_2(t) \end{bmatrix} = \begin{bmatrix} 0 & \kappa(t) \\ -\kappa(t) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{e}_1(t) \\ \mathbf{e}_2(t) \end{bmatrix}$$

$$\mathbf{3\ dimensions:} \quad \begin{bmatrix} \mathbf{e}'_1(t) \\ \mathbf{e}'_2(t) \\ \mathbf{e}'_3(t) \end{bmatrix} = \begin{bmatrix} 0 & \kappa(t) & 0 \\ -\kappa(t) & 0 & \tau(t) \\ 0 & -\tau(t) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{e}_1(t) \\ \mathbf{e}_2(t) \\ \mathbf{e}_3(t) \end{bmatrix}$$

$$\mathbf{n\ dimensions:} \quad \begin{bmatrix} \mathbf{e}'_1(t) \\ \vdots \\ \mathbf{e}'_n(t) \end{bmatrix} = \begin{bmatrix} 0 & \chi_1(t) & & & 0 \\ -\chi_1(t) & \ddots & \ddots & & \\ & \ddots & 0 & \chi_{n-1}(t) & \\ 0 & & -\chi_{n-1}(t) & 0 & \end{bmatrix} \begin{bmatrix} \mathbf{e}_1(t) \\ \vdots \\ \mathbf{e}_n(t) \end{bmatrix}$$

3.3 Curvatura di una superficie, Prima Forma Fondamentale

Su una superficie bidimensionale la curvatura varia a seconda della direzione in cui viene calcolata; è tuttavia possibile darne una misura a partire solo da alcune direzioni particolarmente significative.

Dato un punto P di una superficie Σ , si considerano tutti i piani passanti per la normale a Σ in P : l'intersezione di ogni piano con la superficie determina una curva piana di cui è possibile calcolare la curvatura tramite le già esposte formule di Frenet-Serret con la convenzione che la curvatura sarà positiva se la curva devia nello stesso verso della normale, negativa nel caso opposto. I valori massimi e minimi così ottenuti sono detti **curvature principali**, e le rispettive direzioni anch'esse principali. La determinazione di tali direzioni può essere effettuata tramite l'**operatore forma** o di **Weingarten** che definiremo a seguire. Note le due curvature principali κ_1 e κ_2 , è possibile definire due diverse misure della curvatura, la **curvatura gaussiana** data dal prodotto delle due curvature principali, e la **curvatura media**, che ne è, come suggerisce il nome, la media aritmetica.

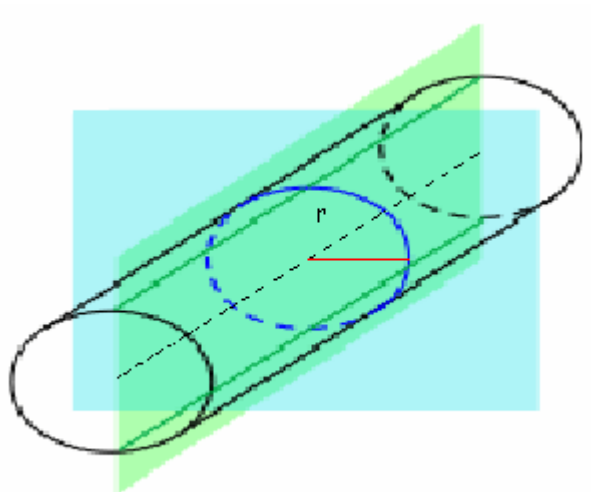
Karl F. Gauss dimostrò nel suo *Theorema Egregium* che la curvatura gaussiana non dipende dall'immersione della superficie in uno spazio ambiente più grande, e può essere definita utilizzando solo le caratteristiche della superficie stessa, si tratta quindi una curvatura *intrinseca*. La curvatura media invece dipende dallo spazio di immersione ed è pertanto una misura di curvatura *estrinseca*.

Prima di approfondire la trattazione matematica in modo più dettagliato vediamo un semplice esempio di calcolo di curvatura secondo il principio di individuazione delle curvature principali.

Curvatura della superficie cilindrica

Consideriamo un cilindro di raggio r , è semplice verificare che le due direzioni principali sono quella parallela e perpendicolare all'asse del cilindro, le cui

rispettive curvatures, per quanto già detto in precedenza, valgono $\kappa_1 = 0$ (si tratta della curvatura di una retta) e $\kappa_2 = 1/r$ (è infatti la curvatura di una circonferenza).



La curvatura gaussiana del cilindro vale dunque $K = \kappa_1 \kappa_2 = 0 \cdot 1/r = 0$, pertanto la geometria del cilindro è piatta (come avevamo già in precedenza affermato ed ora verifichiamo), mentre la curvatura media è pari a $H = (\kappa_1 + \kappa_2/2) = (0 + 1/r/2) = 1/2r$.

Il linguaggio moderno della Geometria Differenziale è quello della geometria e dell'algebra lineare, si parla di applicazioni e di operatori, ad esempio l'ente matematico tensore che abbiamo definito nel secondo capitolo "per componenti" viene visto come un'applicazione multilineare tra spazi vettoriali. Questo linguaggio consente di trattare in modo più agile gli argomenti che andremo ad esporre ora, e pertanto lo adotteremo; una tale visuale consente inoltre di sfruttare i risultati noti dall'algebra lineare per lavorare con gli strumenti di cui ci serviremo. Diamo ora una lettura in termini di questo linguaggio dei concetti già introdotti riguardanti la metrica, e vediamo come si legano ai concetti di curvatura. Le definizioni più importanti per l'obiettivo del lavoro sono riportate, per tutte le altre si rimanda agli opportuni riferimenti posti in bibliografia.

Lemma 3.3.1

Sia $f : U \rightarrow \mathbb{R}$ una funzione differenziabile di un insieme aperto $U \subset \mathbb{R}^2$. Allora il grafico di f , ovvero l'insieme

$$\{(u_1, u_2, f(u_1, u_2)) : (u_1, u_2) \in U\} \subset \mathbb{R}^3$$

è una **superficie differenziabile regolare**. Una superficie Σ espressa in tal modo si dice essere in forma esplicita.

Lemma 3.3.2

Sia $a = (u_1(t_0), u_2(t_0))$ un punto regolare per Σ , consideriamo l'insieme delle curve α su Σ passanti per a , cioè $\alpha(t_0) = X(a)$ (e dove X è una parametrizzazione regolare). Allora la totalità di tutti i vettori tangenti in quel punto forma uno spazio bidimensionale. Tale spazio viene denominato **piano tangente** a Σ in a ed indicato con $T_a(\Sigma)$, ed è uno spazio vettoriale.

Dunque una superficie regolare ammette in ogni punto regolare un piano tangente che risulta, per definizione, indipendente dai parametri (il punto è sempre a , quale che sia la parametrizzazione scelta).

Definizione 3.3.1

Sia Σ una superficie differenziabile, $p \in \Sigma$ e $\mathbf{v} \in T_p(\Sigma)$. Allora esiste una curva $\alpha(t) = X(u_1(t), u_2(t))$, con X parametrizzazione regolare e $t \in [-\varepsilon, \varepsilon]$, tale che

$$\begin{cases} \alpha(0) = p \\ \alpha'(0) = \mathbf{v} \end{cases}$$

α è detta **curva adattata** a \mathbf{v} .

Ad ogni parametrizzazione regolare X di una superficie differenziabile $\Sigma \subset \mathbb{R}^3$ è possibile associare la seguente terna di campi vettoriali differenziabili:

$$\begin{aligned}\mathbf{X}_1 &= \frac{\partial X}{\partial u_1} = \left(\frac{\partial x_1}{\partial u_1}, \frac{\partial x_2}{\partial u_1}, \frac{\partial x_3}{\partial u_1} \right) \\ \mathbf{X}_2 &= \frac{\partial X}{\partial u_2} = \left(\frac{\partial x_1}{\partial u_2}, \frac{\partial x_2}{\partial u_2}, \frac{\partial x_3}{\partial u_2} \right) \\ \mathbf{n} &= \frac{\mathbf{X}_1 \wedge \mathbf{X}_2}{\|\mathbf{X}_1 \wedge \mathbf{X}_2\|}\end{aligned}$$

Osservazioni

1. \mathbf{X}_1 e \mathbf{X}_2 dipendono dalla parametrizzazione X
2. $\forall p \in \Sigma$ $\{\mathbf{X}_1(p), \mathbf{X}_2(p)\}$ formano una base per $T_p(\Sigma)$, in generale non ortonormale
3. \mathbf{n} così definito è detto *campo di versori normali lungo Σ* ed è univocamente determinato. Inoltre dipende solo dall'orientazione scelta su Σ , un'orientazione opposta determina un versore opposto (si ricordi quanto detto prima a proposito del verso del vettore normale).

Prima Forma Fondamentale

Il naturale prodotto interno in \mathbb{R}^3 induce su ogni spazio tangente $T_p(\Sigma)$ di una superficie regolare, un prodotto interno g_p così definito:

$$\begin{aligned}g_p : T_p(\Sigma) \times T_p(\Sigma) &\rightarrow \mathbb{R} \\ (\mathbf{v}, \mathbf{w}) &\mapsto g_p(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle_p = \mathbf{v} \cdot \mathbf{w}\end{aligned}$$

dove $(-, -)$ è l'usuale prodotto scalare in \mathbb{R}^3 e $\mathbf{v}, \mathbf{w} \in T_p(\Sigma)$. Tale prodotto interno è simmetrico e lineare sia rispetto a \mathbf{v} che rispetto a \mathbf{w} , si dice infatti essere **bilineare**.

Assegnare un prodotto scalare su ogni spazio tangente equivale a dotare la superficie Σ di una **metrica riemanniana** g e la coppia (Σ, g) è detta **superficie riemanniana**.

A g_p possiamo associare una forma quadratica definita positiva

$$I_p : T_p(\Sigma) \rightarrow \mathbb{R}$$

$$\mathbf{w} \mapsto I_p(\mathbf{w}) = \langle \mathbf{w}, \mathbf{w} \rangle_p = \|\mathbf{w}\|^2 \geq 0$$

e dove $\|\mathbf{w}\|$ è la lunghezza (modulo) di $\mathbf{w} \in T_p(\Sigma)$.

Definizione 3.3.2

La forma quadratica I_p su $T_p(\Sigma)$ è detta **Prima Forma Fondamentale** della superficie regolare $\Sigma \subset \mathbb{R}^3$ in $p \in \Sigma$.

È possibile esprimere I_p in termini della base $\{\mathbf{X}_1, \mathbf{X}_2\}$ di $T_p(\Sigma)$ associata alla parametrizzazione X scelta.

Sia $\mathbf{w} \in T_p(\Sigma)$ e $\alpha(t) = X(u_1(t), u_2(t))$ con $t \in [-\varepsilon, \varepsilon]$ curva su Σ adattata a \mathbf{w} , tale quindi che $\alpha(0) = p$ e $\mathbf{w} = \alpha'(0) = \mathbf{X}_1(u_1(0), u_2(0))u_1'(0) + \mathbf{X}_2(u_1(0), u_2(0))u_2'(0)$. Allora

$$\begin{aligned} I_p(\mathbf{w}) &= I_p(\alpha'(0)) = \langle \alpha'(0), \alpha'(0) \rangle_p = \\ &= \langle \mathbf{X}_1 u_1' + \mathbf{X}_2 u_2', \mathbf{X}_1 u_1' + \mathbf{X}_2 u_2' \rangle_p = \\ &= \langle \mathbf{X}_1, \mathbf{X}_1 \rangle_p (u_1')^2 + 2 \langle \mathbf{X}_1, \mathbf{X}_2 \rangle_p u_1' u_2' + \langle \mathbf{X}_2, \mathbf{X}_2 \rangle_p (u_2')^2 = \\ &= g_{ij} u_i' u_j' \end{aligned}$$

dove è sottintesa la sommatoria rispetto agli indici ripetuti nell'ultima espressione, tutte le espressioni sono valutate in $t = 0$ e le quantità

$$\begin{aligned}
g_{11} &= E = \langle \mathbf{X}_1, \mathbf{X}_1 \rangle_p \\
g_{12} = g_{21} &= F = \langle \mathbf{X}_1, \mathbf{X}_2 \rangle_p \\
g_{22} &= G = \langle \mathbf{X}_2, \mathbf{X}_2 \rangle_p
\end{aligned}$$

sono i coefficienti della prima forma fondamentale rispetto alla base $\{\mathbf{X}_1(p), \mathbf{X}_2(p)\}$ di $T_p(\Sigma)$. Detta \mathcal{G} la matrice di elementi g_{ij} , vale la relazione $\mathcal{G} = J'J$, dove J è lo Jacobiano. Ritroviamo quindi il tensore metrico così come l'avevamo definito nel paragrafo 2.7, dunque esso non è altro che la matrice associata alla Prima Forma Fondamentale, che è dunque l'applicazione che definisce la metrica. Vale inoltre anche la relazione $\det(\mathcal{G}) = \|\mathbf{X}_1 \wedge \mathbf{X}_2\|^2$.

I_p ci consente di effettuare delle misure sulla superficie (lunghezza di una curva, angolo fra due vettori tangenti, area di una regione..) senza dover fare riferimento allo spazio ambiente \mathbb{R}^3 in cui si trova la superficie.

Come esempio interessante di questa possibilità mostriamo l'espressione per il calcolo dell'angolo θ sotto cui due curve regolari $\alpha: I \rightarrow S, \beta: I \rightarrow S$ si incontrano in $t = t_0$:

$$\cos \theta = \frac{\langle \alpha'(t_0), \beta'(t_0) \rangle}{\|\alpha'(t_0)\| \|\beta'(t_0)\|}$$

e considerando in particolare l'angolo φ fra le curve coordinate C_1, C_2 di una parametrizzazione X in un punto (u_1, u_2) è

$$\cos \varphi = \frac{\langle \mathbf{X}_1, \mathbf{X}_2 \rangle}{\|\mathbf{X}_1\| \|\mathbf{X}_2\|} = \frac{F}{\sqrt{EG}}$$

ne segue che le curve coordinate di una parametrizzazione sono ortogonali se e solo se $F(u_1, u_2) = 0 \quad \forall (u_1, u_2)$. Simili parametrizzazioni sono dette, appunto, ortogonali.

3.4 Operatore di Weingarten e Seconda Forma Fondamentale

Definizione 3.4.1

Sia $\Sigma \subset \mathbb{R}^3$ una superficie con orientazione \mathbf{n} . L'applicazione differenziabile

$$\begin{aligned} N : \Sigma &\rightarrow \Sigma^2 \\ p &\mapsto \mathbf{n}(p) \end{aligned}$$

dove $\mathbf{n}(p)$ è il versore normale a Σ in p e Σ^2 è la sfera unitaria:

$$\Sigma^2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$$

è detta **superficie normale** o **mappa di Gauss** di Σ .

Osservazione

La mappa di Gauss N mette in relazione la geometria di Σ con quella della sfera Σ^2 ; in generale N dà informazioni sulla geometria locale di Σ e su quanto questa si discosti dall'essere una sfera (intuitivamente è la generalizzazione dell'idea di quanto una curva si discosti dal cerchio osculatore in un suo punto).

Consideriamo il differenziale della mappa di Gauss:

$$d_p N : T_p(\Sigma) \rightarrow T_{\mathbf{n}(p)}(\Sigma^2),$$

poiché $\forall p \in \Sigma$ si ha

$$T_{\mathbf{n}(p)}(\Sigma^2) = \mathbf{n}(p)^\perp = T_p(\Sigma)$$

$d_p N$ può essere visto come un operatore su $T_p(\Sigma)$:

$$d_p N : T_p(\Sigma) \rightarrow T_p(\Sigma)$$

$$\mathbf{v} \mapsto d_p N(\mathbf{v}) = \nabla_{\mathbf{v}}^{\mathbb{R}^3} \mathbf{n} \in T_p(\Sigma)$$

Definizione 3.4.2

Sia $\Sigma \subset \mathbb{R}^3$ una superficie orientata. $\forall p \in \Sigma$ definiamo **operatore forma** o **operatore di Weingarten** l'operatore lineare W_p così definito

$$W_p : T_p(\Sigma) \rightarrow T_p(\Sigma)$$

$$\mathbf{v} \mapsto -\nabla_{\mathbf{v}}^{\mathbb{R}^3} \mathbf{n} \quad \forall \mathbf{v} \in T_p(\Sigma)$$

La linearità dell'operatore forma è una conseguenza della linearità della derivata rispetto al vettore per cui si deriva. La terminologia usata è giustificata dal fatto che W_p descrive la variazione infinitesima del versore normale nella direzione di \mathbf{v} , e tale variazione dipende dalla forma di Σ in un intorno del punto p .

Osservazione

Dalla definizione risulta chiaro che l'operatore forma dipende solo dal versore normale \mathbf{n} . Quindi un'orientazione opposta su Σ determina un operatore forma di segno opposto.

Preso α curva adattata a \mathbf{v} , $\alpha'(0) = \mathbf{X}_1 u_1'(0) + \mathbf{X}_2 u_2'(0)$.

$$\nabla_{\mathbf{v}}^{\mathbb{R}^3} \mathbf{n} = \mathbf{n}(u_1'(0), u_2'(0)) = \mathbf{n}_1 u_1' + \mathbf{n}_2 u_2'$$

dove $\mathbf{n}_i = \nabla_{\mathbf{x}_i}^{\mathbb{R}^3} \mathbf{n} \in T_p(\Sigma)$.

Poiché $\mathbf{n}_i \in T_p(\Sigma)$ possiamo scrivere

$$\begin{aligned}\mathbf{n}_1 &= -a_{11}\mathbf{X}_1 - a_{12}\mathbf{X}_2 \\ \mathbf{n}_2 &= -a_{21}\mathbf{X}_1 - a_{22}\mathbf{X}_2\end{aligned}$$

e pertanto

$$\nabla_{\mathbf{v}}^{\mathbb{R}^3} \mathbf{n} = -(a_{11}u'_1 + a_{12}u'_2)\mathbf{X}_1 - (a_{21}u'_1 + a_{22}u'_2)\mathbf{X}_2$$

dove $\forall p \in \Sigma$

$$\mathcal{L} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

è la matrice associata all'operatore forma W_p rispetto alla base $\{\mathbf{X}_1, \mathbf{X}_2\}$ di $T_p(\Sigma)$.

Si può dimostrare che la curvatura media e la curvatura gaussiana, così come la abbiamo definite in precedenza, possono essere espresse in funzione delle componenti di questa matrice come segue:

Definizione 3.4.3

$\forall p \in \Sigma$ definiamo la quantità $H(p) = \frac{1}{2} \text{tr}(\mathcal{L})$ **curvatura media** di Σ in p , e

$K(p) = \det(\mathcal{L})$ **curvatura gaussiana** di Σ in p .

Un'importante proprietà dell'operatore forma è che è simmetrico rispetto al prodotto interno g_p , cioè si ha

$$g_p(W_p(\mathbf{v}), \mathbf{w}) = g_p(\mathbf{v}, W_p(\mathbf{w})) \quad \forall \mathbf{v}, \mathbf{w} \in T_p(\Sigma)$$

Ciò ci consente di applicare il teorema spettrale e di affermare quanto segue:

Teorema 3.4.1

Esiste una base ortonormale $\{\mathbf{T}_1, \mathbf{T}_2\}$ di $T_p(\Sigma)$ rispetto alla quale la matrice associata all'operatore W_p ha la forma

$$\mathcal{L}_{\{\mathbf{T}_1, \mathbf{T}_2\}} = \begin{bmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{bmatrix}$$

dove κ_1 e κ_2 sono gli autovalori di W_p .

Definizione 3.4.4

$\forall p \in \Sigma$ $\kappa_1(p)$ e $\kappa_2(p)$ sono dette **curvature principali** di Σ in p e i relativi autospazi individuati dai versori $\mathbf{T}_1(p)$ e $\mathbf{T}_2(p)$ detti **direzioni principali**.

Osservazione

- $\mathbf{T}_1(p)$ e $\mathbf{T}_2(p)$ sono univocamente determinati a meno del segno e dell'ordine se $\kappa_1(p) \neq \kappa_2(p)$
- $\kappa_1, \kappa_2 : \Sigma \rightarrow \mathbb{R}$ sono funzioni continue e differenziabili su $\Sigma \setminus \{\kappa_1(p) = \kappa_2(p)\}$; un'orientazione opposta su Σ determina curvature opposte.

In questi termini, rispetto alla base $\{\mathbf{T}_1, \mathbf{T}_2\}$ si ha $\forall p \in \Sigma$

$$H(p) = \frac{1}{2}(\kappa_1(p) + \kappa_2(p))$$

$$K(p) = \kappa_1(p)\kappa_2(p)$$

queste ultime possono essere riscritte come

$$\kappa^2 - 2H\kappa + K = 0$$

equazione di secondo grado le cui soluzioni sono

$$\kappa_1 = H + \sqrt{H - K}$$

$$\kappa_2 = H - \sqrt{H - K}$$

Possiamo associare all'operatore forma W_p una forma bilineare II_p su $T_p(\Sigma)$ così definita:

$$II_p : T_p(\Sigma) \times T_p(\Sigma) \rightarrow T_p(\Sigma)$$

$$(\mathbf{v}, \mathbf{w}) \mapsto II_p(\mathbf{v}, \mathbf{w}) = g_p(W_p(\mathbf{v}), \mathbf{w})$$

II_p è detta **Seconda Forma Fondamentale** di Σ in p .

Osservazioni

Anche II_p , come W_p , dipende dal segno di $\mathbf{n}(p)$ e quindi dalla parametrizzazione X . Inoltre $\forall p \in \Sigma$ la Seconda Forma Fondamentale è una forma bilineare simmetrica su $T_p(\Sigma)$ (è una conseguenza del fatto che W_p è simmetrico).

Rispetto alla base $\{\mathbf{X}_1(p), \mathbf{X}_2(p)\}$ di $T_p(\Sigma)$ la matrice associata alla Seconda Forma Fondamentale è

$$\mathcal{B} = \begin{bmatrix} e & f \\ f & g \end{bmatrix}$$

con

$$\begin{aligned} b_{11} = e &= II_p(\mathbf{X}_1(p), \mathbf{X}_1(p)) = \langle -\nabla_{\mathbf{x}_1} \mathbf{n}, \mathbf{X}_1 \rangle_p = \langle \mathbf{n}, \mathbf{X}_{11} \rangle_p \\ b_{12} = b_{21} = f &= II_p(\mathbf{X}_1(p), \mathbf{X}_2(p)) = \langle -\nabla_{\mathbf{x}_1} \mathbf{n}, \mathbf{X}_2 \rangle_p = \langle \mathbf{n}, \mathbf{X}_{12} \rangle_p \\ b_{22} = g &= II_p(\mathbf{X}_2(p), \mathbf{X}_2(p)) = \langle -\nabla_{\mathbf{x}_2} \mathbf{n}, \mathbf{X}_2 \rangle_p = \langle \mathbf{n}, \mathbf{X}_{22} \rangle_p \end{aligned}$$

Dalla definizione di II_p si ha:

$$\begin{bmatrix} e & f \\ f & g \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \begin{bmatrix} E & F \\ F & G \end{bmatrix}$$

quindi

$$\mathcal{L} = \mathcal{G}^{-1} \mathcal{B}$$

dove

$$\mathcal{G}^{-1} = [g^{ij}] = \frac{1}{EG - F^2} \begin{bmatrix} G & -F \\ -F & E \end{bmatrix}$$

Pertanto, in termini dei coefficienti di Prima e Seconda Forma Fondamentale, si ha

$$\begin{aligned} H(p) &= \frac{1}{2} \text{tr}(\mathcal{L}) = \frac{1}{2} \frac{Eg + Ge - 2Ff}{EG - F^2} = \frac{1}{2} g^{ij} b_{ij} \\ K(p) &= \det(\mathcal{L}) = \frac{eg - f^2}{EG - F^2} = \frac{\det(\mathcal{B})}{\det(\mathcal{G})} \end{aligned}$$

e dove è sottintesa la sommatoria rispetto agli indici ripetuti.

Siamo così in grado di esprimere le misure di curvatura, introdotte solo intuitivamente all'inizio del capitolo, in modo rigoroso ed in dipendenza dalle proprietà metriche della superficie, tutte le informazioni a riguardo delle quali sono raccolte nelle due forme fondamentali presentate.

Tutte i concetti presentati per la curvatura, prima di una linea e poi di una superficie, possono essere estese a varietà riemanniane di dimensione $k > 2$ con l'introduzione del **tensore di curvatura di Riemann**, un ente matematico strettamente legato ad una definizione intrinseca di curvatura denominata **trasporto parallelo**: esso si esegue trascinando punto per punto un vettore lungo un percorso chiuso contenuto nella varietà, in modo che la direzione del vettore (riferita alla varietà, e non allo spazio che la contiene) non cambi. Dopo aver percorso un giro completo, il vettore non coincide più con il vettore originario, ma risulta deviato di una quantità che dipende sia dall'area della superficie delimitata dal percorso chiuso, sia dalla curvatura intrinseca della superficie stessa. Risulta evidente da questa presentazione intuitiva che tale deviazione sarà nulla per superfici piate.

Non risulta necessaria nell'ambito di questo lavoro l'esplicitazione formale degli strumenti matematici relativi a tale estensione, in quanto risultano sufficienti le nozioni già introdotte per la comprensione di quanto segue.

Nell'ultima parte della tesi verranno presentati casi immediatamente visualizzabili graficamente, ovvero quelli della curvatura di una linea e di una superficie, cioè di varietà riemanniane di dimensione uno e due che, considerando il parallelismo presentato nel paragrafo 2.8, significa riferirsi a modelli non lineari ad uno o due parametri, e modelli ad un numero maggiore di parametri, per i quali non sarà però ovviamente possibile fornire una rappresentazione grafica soddisfacente delle varietà risultanti.

CAPITOLO IV

La curvatura nei modelli di regressione non lineari nei parametri

In questo capitolo vedremo come le nozioni presentate nei precedenti possano essere sfruttate nell'ambito dell'analisi dei modelli di regressione non lineari nei parametri; si vedrà una scomposizione del cosiddetto *vettore curvatura* che ne individua due componenti, una dipendente dalla parametrizzazione scelta ed una indipendente. Sulla base di tale scomposizione verranno quindi definite le misure di non linearità presentate da Bates e Watts nel loro fondamentale articolo "Relative Curvature Measures of Nonlinearity" apparso sul *Journal of the Royal Statistical Society* nel 1980, di cui segue una libera traduzione (dei paragrafi 1.1, 2.1-2.5, 2.6.1 dell'articolo originale) commentata ed integrata ad opera dell'autore di questa tesi. Si procederà infine ad una analisi delle proprietà degli stimatori ottenuti secondo il criterio dei LS per alcuni modelli ricorrendo a simulazioni.

L'approccio di Bates e Watts sfrutta in modo diretto la misura di curvatura normale e la prima e seconda forma fondamentale, presentate nel terzo capitolo, ma secondo relazioni non esplicitate (come si dirà in seguito, sembra che nella tesi di dottorato del prof. Bates sia stato portato uno studio più approfondito e formalizzato in termini esplicitamente geometrico-differenziale; non è stato però possibile avere accesso a questo documento). Si cita come pienamente aderente all'ottica geometrico-differenziale per l'analisi dei modelli non lineari il lavoro del prof. A.Zanella "Sulle regioni di confidenza per i parametri dei modelli non lineari" (due parti, pubblicate sulla rivista CALCOLO nel 1974 pagine 365-401 e nel 1975 pagine 1-37); il sommario dell'articolo è il seguente:

"Nell'ambito della ricerca sperimentale si presenta il problema di stimare i parametri di un modello interpretativo teorico che ne dipende in modo non lineare. In concreto però, come per il caso lineare, le stime puntuali, ottenute tipicamente secondo il principio dei minimi quadrati, possono avere, in sé, scarso rilievo per la

variabilità ad esse inerente a causa dell'errore sperimentale e vengono impiegate per la determinazione di insiemi di valori che hanno una prefissata probabilità di comprendere quelli degli incogniti parametri in studio (regioni di confidenza).

Nel presente studio si propone un miglioramento, in modo da tener conto della "curvatura", del metodo di uso corrente, che fa ricorso alla "linearizzazione locale" del modello per ottenere regioni di confidenza approssimate. La teoria qui proposta si basa: 1) sull'interpretazione di un modello non lineare in k ($k > 1$) parametri come una varietà riemanniana $V_{(k)}$ immersa in uno spazio euclideo $S_{(n)}$ di dimensioni pari al numero n ($n > k + 1$) delle osservazioni; 2) sulla sua identificazione con una ipersuperficie che, sotto particolari condizioni, meglio approssima $V_{(k)}$, per quanto riguarda la curvatura globale media, nell'intorno del punto corrispondente alle stime dei parametri.

Si propongono due tipi di regioni di confidenza, uno generale ed uno sotto l'ipotesi che la *ipersuperficie* approssimante sia convessa.

La teoria proposta si lascia agevolmente formalizzare in una procedura di calcolo automatico di cui si indicano gli elementi costitutivi e si illustra l'applicazione ad un modello di cinetica chimica"

In questo lavoro viene fatto uso del tensore di curvatura di Riemann e di una sua contrazione, ai fini della quantificazione della curvatura; l'approccio è quindi direttamente collegato alle misure di Geometria Differenziale.

Si è scelto di analizzare il lavoro di Bates e Watts per il carattere più accessibile per chi non abbia solide basi di Geometria Differenziale e Calcolo Tensoriale, e per le interessanti conclusioni, verificate empiricamente tramite le simulazioni presentate nell'ultima parte di questo lavoro, cui è possibile giungere.

Si cita inoltre l'esistenza di un filone di ricerca, che prese l'avvio nei primi anni 80, dedito all'analisi delle proprietà asintotiche delle stime per i modelli non lineari esclusivamente basato sul calcolo tensoriale. Uno dei precursori in questo campo fu Shun-ichi Amari, che pubblicò nel 1984 il volume "Differential-Geometrical Methods in Statistics", edito da Springer-Verlag per la serie "Lectures Notes in

Statistics”; questo ottimo testo presenta come unica difficoltà di lettura la necessità di solide basi relativamente agli strumenti propri del Calcolo Tensoriale affinché possa essere compreso a fondo.

4.1 Scomposizione del vettore curvatura

Sia $\Sigma \subset \mathbb{R}^3$ una superficie differenziabile regolare orientata con parametrizzazione X e $C \subset \Sigma$ una curva differenziabile regolare orientata con parametrizzazione naturale (o parametrizzazione per la lunghezza d’arco) $\alpha(s) = X(u_1(s), u_2(s)), s \in (-\varepsilon, \varepsilon)$.

Il versore tangente alla curva è dato da

$$\mathbf{T}(s) = \alpha'(s) = \frac{dX}{ds} = \frac{\partial X}{\partial u_1} \frac{du_1}{ds} + \frac{\partial X}{\partial u_2} \frac{du_2}{ds}$$

da cui, derivando ancora e tenendo conto delle regole della derivazione composta, otteniamo il **vettore curvatura**:

$$\begin{aligned} \mathbf{K}_C(s) = \alpha''(s) &= \frac{d^2 X}{ds^2} = \left[\frac{\partial X}{\partial u_1} \frac{d^2 u_1}{ds^2} + \frac{\partial X}{\partial u_2} \frac{d^2 u_2}{ds^2} \right] + \left[\frac{\partial^2 X}{\partial u_1^2} \left(\frac{du_1}{ds} \right)^2 + \frac{\partial^2 X}{\partial u_2^2} \left(\frac{du_2}{ds} \right)^2 + 2 \frac{\partial^2 X}{\partial u_1 \partial u_2} \frac{du_1}{ds} \frac{du_2}{ds} \right] = \\ &= \sum_{i=1}^2 \frac{\partial X}{\partial u_i} \frac{d^2 u_i}{ds^2} + \sum_{i,j=1}^2 \frac{\partial^2 X}{\partial u_i \partial u_j} \frac{du_i}{ds} \frac{du_j}{ds} = \mathbf{K}_C^\Sigma(s) + \mathbf{K}_C^N(s) \end{aligned}$$

Tale vettore riassume le informazioni di curvatura (si noti il parallelismo con le componenti di prima e seconda forma fondamentale) ed è composto da due elementi, $\mathbf{K}_C^\Sigma(s)$ e $\mathbf{K}_C^N(s)$.

$\mathbf{K}_C^\Sigma(s)$ è la componente sulla superficie, mentre $\mathbf{K}_C^N(s)$ è la componente lungo la normale alla superficie. Per ogni punto $p \in \Sigma$ si definiscono le seguenti quantità:

$$k_{g,p} = \pm \|\mathbf{K}_{C,p}^\Sigma\| \quad \text{curvatura geodetica di } C \in \Sigma \text{ in } p$$

$$k_{n,p} = \pm \|\mathbf{K}_{C,p}^N\| \quad \text{curvatura normale di } C \in \Sigma \text{ in } p$$

il segno di $k_{g,p}$ dipende dall'orientazione di C e Σ , quello di $k_{n,p}$ solo dall'orientazione di Σ .

Dalle formule di Frenet-Serret si ha

$$\mathbf{K}_C(s) = \alpha''(s) = \kappa(s)\mathbf{N}$$

dove \mathbf{N} è la normale principale alla curva, detto φ l'angolo fra \mathbf{N} (normale principale alla curva) e la normale alla superficie \mathbf{n} si ha

$$k_n = \kappa \langle \mathbf{n}, \mathbf{N} \rangle = \kappa \cos \varphi$$

$$k_g = \kappa \sin \varphi$$

La curvatura geodetica sarà quindi nulla qualora la normale principale della curva e la normale alla superficie coincidano.

Prendendo le mosse da questa scomposizione Bates e Watts introdussero le loro misure relative di non linearità come due quantità riguardanti rispettivamente la natura non lineare **intrinseca** del modello e la non linearità **indotta dai parametri**, considerando una estensione generica a p dimensioni della varietà riemanniana definita dalle medie del modello in funzione dei p parametri.

4.2 Perché introdurre misure di curvatura nell'analisi dei modelli non lineari

Posto che la natura degli errori delle variabili in considerazione in un modello sia casuale (ed essi siano indipendenti e normalmente distribuiti, con media nulla e varianza finita), la significatività della procedura dei LS è dovuta al fatto che le

stime conseguenti dei parametri posseggono alcune proprietà statistiche ottimali. *Esclusivamente per il caso del modello lineare* questo significa che le stime sono non distorte, coincidenti con le stime di ML e raggiungono il limite inferiore di varianza. In questi termini il metodo dei LS fornisce i migliori parametri possibili.

Si tenga a mente che solo stime con le proprietà sopraccitate permettono il calcolo di un affidabile range per la loro varianza e della vera correlazione esistente tra di esse, forniscono previsioni della variabile dipendente con determinata probabilità di errore ed assicurano test statistici significativi (in termini di applicabilità).

Le stime LS per modelli non lineari invece non posseggono le proprietà ottimali di cui si è già discusso. Sono solitamente distorte e distribuite in modo non normale, e la loro varianza eccede in buona misura il minimo di varianza stabilito dal limite inferiore di Rao-Cramér. L'estensione della distorsione, della non normalità e l'eccesso di varianza variano in modo significativo da modello a modello, ed anche all'interno dello stesso modello (fra parametri differenti); ciò dipende da un gran numero di fattori, incluso l'ammontare dei dati disponibili e la loro distribuzione nello spazio sperimentale. È quindi forse più appropriato parlare di una combinazione *dati-modello* nell'analisi della non linearità e delle stime dei modelli non lineari.

È noto che, comunque, le stime LS dei parametri nei modelli non lineari raggiungono le proprietà statistiche ottimali delle stime del modello lineare asintoticamente e quindi, sotto le condizioni che le variabili dipendenti (i *fitted values*) posseggano errori che rispettino le ipotesi già introdotte, più grande sarà il campione, più vicine saranno le proprietà delle stime a quelle ottimali. Sfortunatamente non è possibile dire a priori quanto grande debba essere un campione perché ci si avvicini sufficientemente a queste proprietà asintotiche. Per alcuni modelli non lineari tali proprietà sono bene approssimate anche per piccole dimensioni campionarie (6-10 unità), mentre per altri modelli questo non è vero se non in presenza di numerosità elevate (50-100), in altri casi ancora servono campioni ulteriormente più grandi, che potrebbero non essere materialmente ottenibili (ad esempio per motivi economici, potrebbero essere esperimenti molto costosi).

Fino ad ora non si conosce una regola che stabilisca a priori la dimensione campionaria per ottenere le desiderate proprietà a fronte di un modello non lineare, qualunque miglioramento in merito può essere ottenuto solamente a posteriori, ma per poter far questo è necessario disporre di qualche misura quantitativa di *deviazione dalla linearità* (ovvero di non linearità), che consenta di orientare gli sforzi in modo fruttifero.

In letteratura sono presenti molti tentativi per quantificare questa deviazione, ovvero quanto diversamente il modello non lineare si comporti rispetto a quello lineare, e per fornire una metodologia di valutazione del comportamento *vicino al lineare* (close to linear). Le prime misure quantitative di comportamento non lineare furono proposte nel 1960 da Beale. Guttman e Meeter qualche anno più tardi scrissero a riguardo delle limitazioni del metodo di Beale. Nel 1971 Box derivò alcune formule per stimare la distorsione delle stime secondo i LS, che vennero esaminate successivamente da Gillis e Ratkowski tramite approfonditi studi di simulazione. Contemporaneamente vennero introdotte misure di non linearità da Bates e Watts nel loro fondamentale lavoro del 1980. Essi svilupparono nuove misure di non linearità basate sul concetto intuitivo di curvatura noto dalla geometria differenziale, applicandolo a spazi multi-dimensionali (tali concetti sono quelli che abbiamo introdotto nel corso di questa tesi, ed in particolare nel terzo capitolo); essi fornirono inoltre relazioni tra le loro misure e quelle proposte in precedenza da Beale, e mostrarono come la misura di distorsione di Box fosse collegata alla loro proposta.

Bates e Watts stabilirono che la non linearità di un modello può essere convenientemente separata in due componenti:

- **non linearità intrinseca**, d'ora in poi **IN** (*intrinsic*)
- **non linearità indotta dai parametri**, d'ora in poi **PE** (*parameters-effect*)

Non linearità intrinseca

La non linearità **IN** rappresenta la curvatura della superficie multi-dimensionale chiamata luogo delle soluzioni del modello. Essa è legata al particolare modello ed all'insieme di dati, il numero dei dati e la loro distribuzione nello spazio e non può essere modificata da trasformazioni matematiche dei parametri (riparametrizzazione), risultandone quindi indipendente. Solitamente decresce all'aumentare dei dati per un determinato modello, non è però nota alcuna relazione riguardante la misura in cui questo accada. La curvatura **IN** è pari a zero per il modello lineare, e si avvicinerà a tale valore per i modelli non lineari asintoticamente, ovvero al tendere ad infinito della dimensione campionaria. È interessante sottolineare quanto un valore ridotto di **IN** implichi una distorsione trascurabile nella previsione della variabile dipendente, e renda i limiti di confidenza determinati significativi. Geometricamente indica quanto, in prossimità del punto di stima parametrica, il luogo delle soluzioni si discosti dal suo piano tangente (o dalla sua tangente, nel caso di un solo parametro e quindi di una curva come luogo delle soluzioni).

Non linearità indotta dai parametri

La non linearità **PE** è una misura di curvatura associata alla forma matematica e alla particolare scelta dei parametri, in tal senso può essere modificata tramite la riparametrizzazione. Tramite una opportuna riparametrizzazione, può essere ridotta ad un qualunque valore desiderato (idealmente zero). Geometricamente la curvatura **PE** è collegata alla proiezione delle linee parametriche sul piano tangente al luogo delle soluzioni nel punto di stima: se il modello è lineare questo darà luogo ad una griglia di proiezioni delle linee coordinate parallele ed equispaziate tra di loro, cosa che non avverrà nel caso di un modello non lineare; quanto più la proiezione sarà vicina ad avere le proprietà appena enunciate, tanto più il modello potrà dirsi prossimo alla linearità nel comportamento dei parametri. Nella condizione di una **IN** trascurabile, tanto più piccola sarà la non linearità indotta dai parametri, tanto più il

modello nel suo complesso approssimerà il comportamento di un modello lineare. Di conseguenza sarà necessario un numero inferiore di iterazioni per raggiungere la convergenza nella stima dei parametri secondo l'algoritmo impiegato (da notare che nel caso del modello lineare il calcolo non richiede iterazioni), e formule e test statistici impiegati per il modello lineare (ad esempio il test t e la correlazione dei parametri) saranno da ritenersi più validi. In breve, i parametri stimati saranno più vicini ai loro stimatori di minima varianza.

La misura di distorsione di Box

Tale misura e la *distorsione percentuale* (percentage bias) furono proposte da Box nel 1971. Sono misure di non linearità indirettamente legate ad **IN** e **PE**, e rappresentano entrambe misure individuali di non linearità. Sono collegate al singolo parametro e a come questo compare nel modello, e possono indicare quei parametri che si comportano nel modo "più non lineare", e suggerire quindi verso quali orientare una eventuale riparametrizzazione (le misure di distorsione di Box risentono quindi di trasformazioni parametriche). Possono essere impiegate inoltre per valutare a priori l'impatto di una qualsivoglia riparametrizzazione senza dover stimare nuovamente i parametri del modello, tramite un'opportuna relazione. La distorsione percentuale esprime la distorsione del parametro in relazione alla sua stima secondo i minimi quadrati, è una quantità utile che quantifica in una certa misura il comportamento non lineare del parametro stesso. Escludendo il caso in cui un parametro rappresenti un termine costante nel modello, quanto più vicino il percentage bias sarà allo zero, tanto apparirà più prossimo il suo comportamento a quello lineare. È stata introdotta una regola arbitraria che afferma che una distorsione percentuale superiore all'1% è una buona indicazione di comportamento non lineare; è però importante sottolineare che l'apparire linearmente di un parametro nell'espressione di un modello non lineare non necessariamente implica un comportamento lineare nella stima, un Box bias pari a zero o altre proprietà statistiche che possiamo attenderci da un modello lineare.

Lo Standard Radius

Un importante elemento legato alle misure di curvatura di non linearità ed impiegato nella loro computazione è il *Raggio Standard* (Standard Radius, d'ora in poi **SR**). Si tratta della radice quadrata della somma dei residui al quadrato, divisa per i gradi di libertà e moltiplicata per il numero dei parametri. Viene usata per ponderare opportunamente (vedremo in seguito perché) la curvatura reale del luogo delle soluzioni, sulla base della quale nei calcoli si ottengono le misure di **IN** e **PE**. Inoltre è anche il fattore di riferimento per la definizione di regioni di confidenza cosiddette *conservative*; tali regioni sono costruite di modo che la probabilità che il vero valore del parametro vi cada all'interno sia almeno $1 - \alpha$.

In sintesi, l'impostazione presentata fornisce informazioni riguardo a quanto le proprietà degli stimatori di un modello non lineare in esame siano prossime a quelle riguardanti il modello lineare. Se tale prossimità è sufficiente, allora si potranno ragionevolmente impiegare i test, le formule e i metodi standard dell'inferenza statistica nei confronti dei parametri. Come valutare però tale "sufficiente prossimità"? essa può essere verificata confrontando i valori di **IN** e **PE** con la statistica $1/\sqrt{F}$, dove $F = F(p, n - p; \alpha)$ è il valore ottenibile dalle tavole di una F di Snedecor con p ed $n - p$ gradi di libertà (p è il numero dei parametri del modello ed n la dimensione campionaria) ad un livello di probabilità α . Il valore $1/\sqrt{F}$ può essere visto come il raggio di curvatura della regione di confidenza al $100 \cdot (1 - \alpha)\%$. L'introduzione di queste misure di non linearità ci possono aiutare a valutare l'efficacia di un particolare modello, e facilitano l'eventuale scelta fra modelli diversi.

4.3 Costruzione delle misure di non linearità

Le tecniche di approssimazione nella stima

Prima di procedere al calcolo delle misure di non linearità riprendiamo una questione solo accennata nel paragrafo 1.5, ovvero come sia possibile procedere alla stima dei parametri in un modello non lineare.

Esistono numerosi algoritmi e procedure in letteratura a questo riguardo, alcuni dei più famosi sono la “Marquadt-Levenberg Strategy” (impiegata nella maggior parte dei software commerciale per la stima parametrica) e l’algoritmo “rank-two DFP”, che si distingue principalmente per una alta velocità di convergenza alla soluzione. Entrambi i metodi (così come la quasi totalità di quelli presenti) fanno riferimento alla procedura di Gauss-Newton, ovvero approssimazione della funzione non lineare tramite una espansione troncata al primo ordine in serie di Taylor nell’intorno di un punto prefissato, e differiscono per le correzioni alle stime introdotte iterazione dopo iterazione. Questo porta principalmente a differenze nella velocità della convergenza e nell’affidabilità dell’approssimazione. Metodi di questo genere possono essere riassunti come segue.

Sia fissato un insieme di n combinazioni di livelli dei fattori sperimentali $\mathbf{x}_{1t}, \mathbf{x}_{2t}, \dots, \mathbf{x}_{pt}$, con $t = 1, 2, \dots, n$, e si vogliano stimare i parametri di un modello avente medie della forma $\eta_t = f(\mathbf{x}_{1t}, \mathbf{x}_{2t}, \dots, \mathbf{x}_{pt}; \boldsymbol{\theta})$ (a fronte delle usuali assunzioni sugli errori), dove f è una funzione non lineare dei p incogniti parametri $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$ oggetto di stima. Si puntualizza il fatto che diverse combinazioni di fattori sperimentali determinano differenti varietà riemanniane descritte dalle medie η_t del modello, in conseguenza di questo fatto le considerazioni che seguiranno faranno sempre riferimento ad un generico ma fissato disegno sperimentale.

Una qualunque procedura secondo i LS richiede il calcolo (analitico o numerico) delle derivate parziali prime di η rispetto ad ogni parametro; riprendendo quanto detto nel paragrafo 1.5 riassumiamo brevemente la procedura:

- Nell'intorno di un fissato valore parametrico $\boldsymbol{\theta}_0$ la funzione del modello è approssimata da

$$f(\mathbf{x}, \boldsymbol{\theta}) \cong f(\mathbf{x}, \boldsymbol{\theta}_0) + \sum_{i=1}^p (\mathcal{G}_i - \mathcal{G}_{i,0}) v_i(\mathbf{x})$$

dove

$$v_i(\mathbf{x}) = \partial f(\mathbf{x}, \boldsymbol{\theta}_0) / \partial \mathcal{G}_i |_{\mathcal{G}_0}.$$

- Raggruppando le componenti nei vettori $\boldsymbol{\eta}(\boldsymbol{\theta})$, $\boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$ e \mathbf{v}_i , dove

$$\begin{aligned} \mathbf{v}_i &= (v_i(\mathbf{x}_1), v_i(\mathbf{x}_2), \dots, v_i(\mathbf{x}_n))' = \\ &= \partial \boldsymbol{\eta} / \partial \mathcal{G}_i |_{\mathcal{G}_0}, \quad i = 1, 2, \dots, p, \end{aligned}$$

otteniamo

$$\boldsymbol{\eta}(\boldsymbol{\theta}) \cong \boldsymbol{\eta}(\boldsymbol{\theta}_0) + \sum_{i=1}^p (\mathcal{G}_i - \mathcal{G}_{i,0}) \mathbf{v}_i.$$

- L'effetto di tale approssimazione è di sostituire il luogo delle soluzioni con il suo piano tangente in $\boldsymbol{\eta}(\boldsymbol{\theta}_0)$ (*assunzione planare*), e simultaneamente di imporre un sistema di coordinate uniformi su detto piano. Entrambe l'efficacia degli algoritmi di stima secondo LS e la validità dell'inferenza applicata ai parametri di un modello non lineare saranno fortemente influenzate dalla vicinanza dell'approssimazione lineare al modello.
- La soluzione del problema di LS è quel parametro (vettore parametrico) $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$, denominato appunto *stima secondo i LS*, che minimizza la quantità

$$S(\boldsymbol{\theta}) = \|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\|^2$$

Costruzione degli indici

Il modello fornisce una cosiddetta *mappatura* dallo spazio parametrico p -dimensionale ad una superficie p -dimensionale immersa nello spazio campionario n -dimensionale: come già discusso otterremo quindi una varietà riemanniana di dimensione pari a p immersa in uno spazio euclideo ambiente n -dimensionale. Ciò significa che ogni punto $\boldsymbol{\theta}_0$ nello spazio parametrico mappa ad un punto $\eta(\boldsymbol{\theta}_0)$ sul luogo delle soluzioni; similmente, linee nello spazio parametrico passanti per $\boldsymbol{\theta}_0$ mappano a curve sul luogo delle soluzioni passanti per $\eta(\boldsymbol{\theta}_0)$. Da queste considerazioni prende le mosse il ragionamento che portò Bates e Watts a definire le loro misure di non linearità basandosi sugli strumenti geometrico-differenziali.

Una linea retta arbitraria nello spazio parametrico passante per $\boldsymbol{\theta}_0$ può essere espressa, impiegando il parametro b , come

$$\boldsymbol{\theta}(b) = \boldsymbol{\theta}_0 + b\mathbf{h}$$

e dove $\mathbf{h} = (h_1, h_2, \dots, h_p)'$ è un qualunque vettore non nullo. Questa linea genera una curva $\eta_{\mathbf{h}}(b)$ (chiamata *lifted line*, ovvero una curva adattata, che svolge il ruolo della curva α di cui si è trattato in precedenza) sul luogo delle soluzioni dove

$$\eta_{\mathbf{h}}(b) = \boldsymbol{\eta}(\boldsymbol{\theta}_0 + b\mathbf{h})$$

La tangente alla curva $\eta_{\mathbf{h}}(b)$ in $b=0$ (che ne rappresenta ciò che in fisica viene chiamata **velocità**) è

$$\dot{\eta}_{\mathbf{h}} = \left. \frac{d\eta_{\mathbf{h}}}{db} \right|_{b=0} = \sum_{i=1}^p \left. \frac{\partial \eta}{\partial \theta_i} \right|_{\theta_0} \left. \frac{d\theta_i}{db} \right|_{b=0} = \sum_{i=1}^p \mathbf{v}_i h_i$$

che può essere scritta in forma più compatta impiegando la matrice $\dot{\mathbf{V}}$ di dimensione $n \times p$, la cui i -esima colonna sarà formata dal vettore \mathbf{v}_i già definito, come $\dot{\eta}_{\mathbf{h}} = \dot{\mathbf{V}}\mathbf{h}$.

$$\begin{aligned} \dot{\mathbf{V}}\mathbf{h} &= [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_p] \mathbf{h} = \begin{bmatrix} \frac{\partial \eta_1}{\partial \theta_1} & \dots & \frac{\partial \eta_1}{\partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial \eta_n}{\partial \theta_1} & \dots & \frac{\partial \eta_n}{\partial \theta_p} \end{bmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix} = \begin{pmatrix} h_1 \frac{\partial \eta_1}{\partial \theta_1} + \dots + h_p \frac{\partial \eta_1}{\partial \theta_p} \\ \vdots \\ h_1 \frac{\partial \eta_n}{\partial \theta_1} + \dots + h_p \frac{\partial \eta_n}{\partial \theta_p} \end{pmatrix} = \\ &= h_1 \begin{pmatrix} \frac{\partial \eta_1}{\partial \theta_1} \\ \vdots \\ \frac{\partial \eta_n}{\partial \theta_1} \end{pmatrix} + h_2 \begin{pmatrix} \frac{\partial \eta_1}{\partial \theta_2} \\ \vdots \\ \frac{\partial \eta_n}{\partial \theta_2} \end{pmatrix} + \dots + h_p \begin{pmatrix} \frac{\partial \eta_1}{\partial \theta_p} \\ \vdots \\ \frac{\partial \eta_n}{\partial \theta_p} \end{pmatrix} = \sum_{i=1}^p \mathbf{v}_i h_i \end{aligned}$$

Il vettore tangente ad $\eta_{\mathbf{h}}(b)$ è quindi una combinazione lineare dei vettori \mathbf{v}_i , e l'insieme di tutte tali combinazioni lineari forma il *piano tangente* in $\eta(\theta_0)$.

Nota: Gli algoritmi di stima sfruttano un'approssimazione in serie di Taylor della funzione non lineare troncata al primo ordine, ma per poter ottenere informazioni riguardanti la curvatura sono necessarie informazioni relative alle derivate seconde, nell'ottica di una descrizione più accurata nell'intorno di un qualche punto di riferimento sulla superficie (che nel nostro caso risulta essere il punto di stima parametrica).

L'**accelerazione** della curva $\eta_{\mathbf{h}}$ è anch'essa una combinazione lineare di vettori determinati dai parametri dei modelli e coincide con il vettore delle derivate seconde. Definiamo i vettori delle derivate parziali seconde \mathbf{v}_{ij} come

$$\mathbf{v}_{ij} = \left. \frac{\partial^2 \eta}{\partial \theta_i \partial \theta_j} \right|_{\theta_0}$$

e li raccogliamo in una matrice $p \times p$ $\ddot{\mathbf{V}}$ (come usuale l'indice i individua la riga all'interno di tale matrice e l'indice j la colonna), dove ogni elemento è un n -vettore di componenti $v_{ij} = \partial^2 \eta_t / \partial \theta_i \partial \theta_j$, con $t=1,2,\dots,n$. Il vettore derivata seconda di $\eta_{\mathbf{h}}(b)$ in $b=0$ può quindi essere scritta con riferimento alla sua t -esima componente come

$$\begin{aligned} \ddot{\eta}_{\mathbf{h},t} &= \left. \frac{d^2 \eta_{\mathbf{h}}}{db^2} \right|_0 = \sum_j \left. \frac{\partial \sum_i \mathbf{v}_i h_i}{\partial \theta_j} \right|_{\theta_0} \left. \frac{d\theta_j}{db} \right|_0 = \\ &= \sum_{i,j} v_{ij} h_i h_j = \\ &= \mathbf{h}' \ddot{\mathbf{V}}_t \mathbf{h} \end{aligned}$$

Da questa espressione vediamo quindi che il vettore $\ddot{\eta}_{\mathbf{h}}$ altro non è che una forma quadratica relativa ai vettori \mathbf{v}_{ij} ; $\ddot{\mathbf{V}}$ è un *array* di dimensione $n \times p \times p$, per cui la pre- e post-moltiplicazione per i vettori \mathbf{h}' ($1 \times p$) e \mathbf{h} ($p \times 1$) dà luogo ad un vettore $n \times 1$, ogni elemento del quale è della forma $\mathbf{h}' \ddot{\mathbf{V}}_t \mathbf{h}$, dove $\ddot{\mathbf{V}}_t$ è la t -esima faccia di $\ddot{\mathbf{V}}$

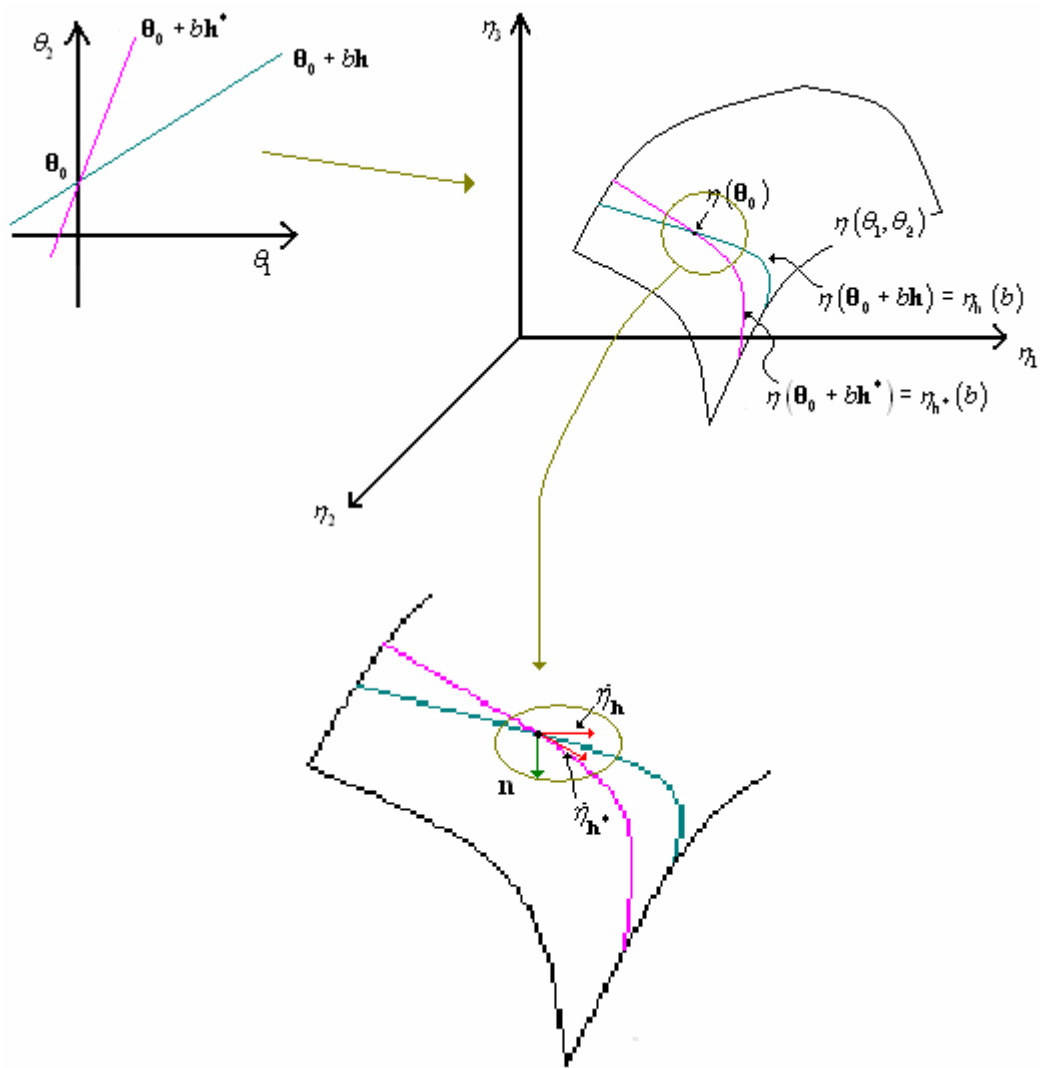
$$\ddot{\mathbf{V}}_t = \begin{bmatrix} \frac{\partial^2 \eta_t}{\partial \theta_1^2} & & \frac{\partial \eta_t}{\partial \theta_1 \partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial \eta_t}{\partial \theta_p \partial \theta_1} & \dots & \frac{\partial \eta_t}{\partial \theta_p^2} \end{bmatrix}_i$$

I vettori $\dot{\eta}_{\mathbf{h}}$ e $\ddot{\eta}_{\mathbf{h}}$ hanno un'interpretazione fisica se immaginiamo un punto che si muove nello spazio campionario di modo che al "tempo" b si trovi in $\eta_{\mathbf{h}}(b)$. La

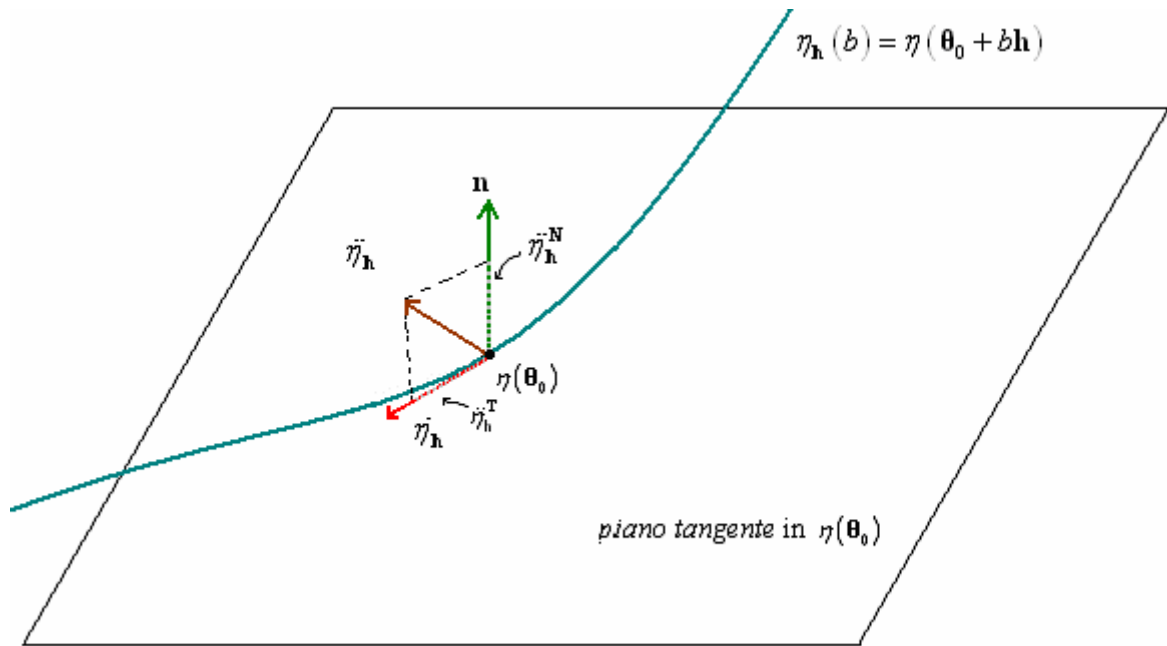
tangente $\dot{\eta}_h$ è la sua **velocità istantanea** e $\ddot{\eta}_h$ l'**accelerazione istantanea** quando $b=0$ (ovvero in corrispondenza del punto θ_0).

Il vettore di accelerazione $\ddot{\eta}_h$, che facciamo notare coincidere con il vettore curvatura introdotto in principio del capitolo, può essere scomposto analogamente a quanto fatto per \mathbf{K}_C . Gli autori propongono una tripartizione di tale vettore nelle componenti $\ddot{\eta}_h^N$ (lungo la normale alla superficie), $\ddot{\eta}_h^P$ e $\ddot{\eta}_h^G$ (componenti *parallela* e *geodetica*, entrambe sul piano tangente); queste ultime due componenti vengono poi sommate per dar luogo alla cosiddetta *accelerazione tangenziale totale*, dunque $\ddot{\eta}_h = \ddot{\eta}_h^N + (\ddot{\eta}_h^P + \ddot{\eta}_h^G) = \ddot{\eta}_h^N + \ddot{\eta}_h^T$.

Ipotizziamo per semplicità di trovarci nel caso $p=2$ (due parametri) e quindi $i, j=1,2$, il grafico nella pagina seguente esplicita il ragionamento presentato nelle pagine precedenti.



Affinché fosse possibile rappresentare graficamente la questione ci si è limitati alla situazione $n=3, p=2$. Al punto θ_0 e alle due linee rette $\theta_0 + b\mathbf{h}$ e $\theta_0 + b\mathbf{h}^*$ nello spazio parametrico corrispondono rispettivamente il punto $\eta(\theta_0)$ e le due curve $\eta_{\mathbf{h}}(b)$ e $\eta_{\mathbf{h}^*}(b)$ sulla varietà descritta dalle medie del modello. I vettori tangenti alle curve nelle direzioni \mathbf{h} e \mathbf{h}^* sono $\dot{\eta}_{\mathbf{h}}$ e $\dot{\eta}_{\mathbf{h}^*}$ (in rosso nel grafico) che come si può vedere hanno direzione diversa rispetto a quella della normale, e questo cambio di direzione è appunto l'informazione contenuta in $\dot{\eta}_{\mathbf{h}}^{\mathbf{N}}$. Il cambio di velocità del punto b è riscontrabile confrontando le componenti tangenti del vettore curvatura nelle due direzioni, e questa informazione è riassunta da $\dot{\eta}_{\mathbf{h}}^{\mathbf{T}}$.



Infine, in marrone, e disegnato soltanto per la direzione \mathbf{h} unicamente per non appesantire il grafico, è visibile il vettore di curvatura, o vettore delle derivate seconde $\ddot{\eta}_{\mathbf{h}}$, con le proiezioni $\ddot{\eta}_{\mathbf{h}}^T$ sul piano tangente e $\ddot{\eta}_{\mathbf{h}}^N$ sulla normale, a mostrare visivamente la sua scomposizione in funzione di questi vettori.

Tornando all'interpretazione fisica, l'accelerazione **normale** $\ddot{\eta}_{\mathbf{h}}^N$ determina la variazione nella *direzione* del vettore $\dot{\eta}_{\mathbf{h}}$ rispetto alla normale al piano tangente al variare della direzione \mathbf{h} , mentre l'accelerazione **tangenziale** $\ddot{\eta}_{\mathbf{h}}^T$ riguarda il cambio di velocità del punto in movimento, e determina dunque se tale punto si muova o meno uniformemente lungo il luogo delle soluzioni.

Bates e Watts propongono la conversione delle componenti di accelerazione in curvatures, ad esempio

$$K_{\mathbf{h}}^N = \frac{\|\ddot{\eta}_{\mathbf{h}}^N\|}{\|\dot{\eta}_{\mathbf{h}}\|^2}$$

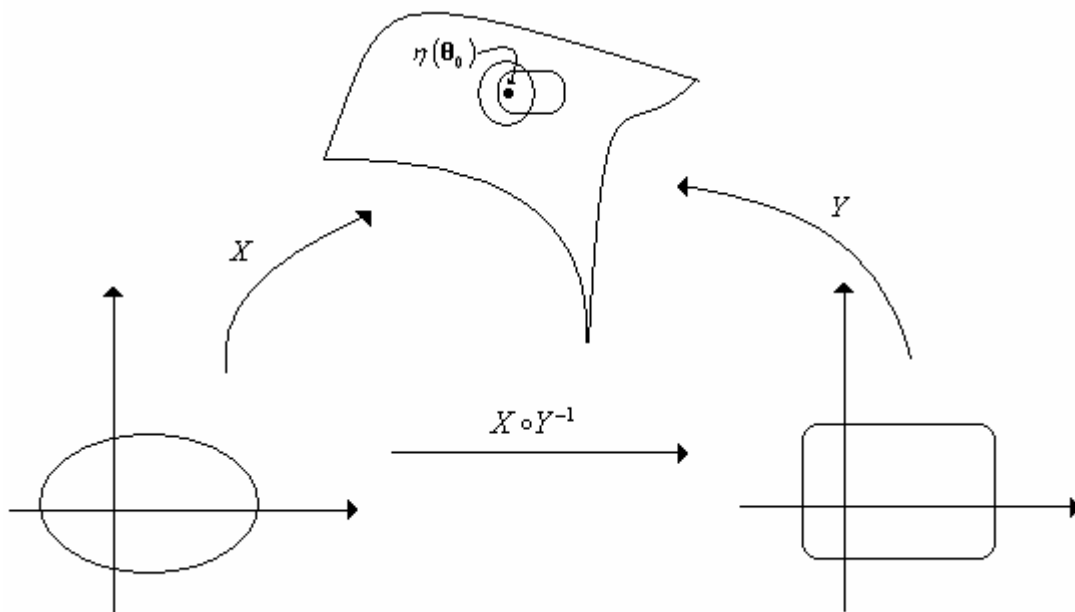
fornisce la curvatura normale nella direzione \mathbf{h} . La sua interpretazione geometrica è dunque quella dell'inverso del raggio del cerchio che meglio approssima il luogo delle soluzioni nella direzione di $\eta_{\mathbf{h}}$ in $\eta(\boldsymbol{\theta}_0)$. La normale ad una superficie è indipendente dalla scelta di parametrizzazione, ovvero dipenderà soltanto dalla combinazione modello matematico-condizioni sperimentali, risulta essere una proprietà intrinseca della superficie. Bates e Watts non dimostrano la coincidenza fra la misura $K_{\mathbf{h}}^N$ e la curvatura normale nota dalla Geometria Differenziale, né l'invarianza della normale rispetto a cambiamenti delle coordinate sulla superficie; questo può essere però verificato come segue, ne diamo dunque dimostrazione per il caso $n=3, p=2$; il risultato è immediatamente generalizzabile a dimensioni superiori.

Vogliamo quindi dimostrare che **la normale alla superficie è indipendente dalla parametrizzazione scelta sulla varietà**. Come già definito in precedenza (paragrafo 3.3) il **campo di vettori normali** alla superficie è (manteniamo le notazioni impiegate in precedenza)

$$\mathbf{n} = \frac{\mathbf{X}_1 \wedge \mathbf{X}_2}{\|\mathbf{X}_1 \wedge \mathbf{X}_2\|}$$

la normale alla superficie sarà dunque un vettore (versore) di questa forma.

Si considerino ora una parametrizzazione regolare $X = X(\theta_1, \theta_2)$ nell'intorno del punto $\eta(\boldsymbol{\theta}_0)$ appartenente al luogo delle soluzioni ed una parametrizzazione alternativa (anch'essa regolare) $Y = Y(\varphi_1(\theta_1, \theta_2), \varphi_2(\theta_1, \theta_2))$ tale per cui $X \circ Y^{-1}$ sia un *diffeomorfismo*, ovvero le funzioni siano invertibili e differenziabili fino all'ordine necessario nell'intorno del punto. Graficamente possiamo riassumere come segue:



Al cambio di coordinate è associata la matrice Jacobiana

$$\mathbf{J} = \frac{\partial X(\theta_1, \theta_2)}{\partial Y(\varphi_1, \varphi_2)}$$

la quale, per le condizioni di regolarità imposte, avrà determinante diverso da zero (il cambiamento di coordinate è invertibile). Consideriamo ora il campo di vettori normali definito nella nuova parametrizzazione da $\mathbf{N}^* = \mathbf{Y}_1 \wedge \mathbf{Y}_2$, dove:

$$\mathbf{Y}_1 = \frac{\partial \Phi}{\partial \varphi_1} \quad \text{e} \quad \mathbf{Y}_2 = \frac{\partial \Phi}{\partial \varphi_2}.$$

In virtù delle regole della derivazione composta avremo che

$$\mathbf{Y}_1 \wedge \mathbf{Y}_2 = \det(\mathbf{J}) \mathbf{X}_1 \wedge \mathbf{X}_2$$

ovvero che gli elementi dei due campi vettoriali sono proporzionali, cioè hanno la stessa direzione; una volta normalizzati a lunghezza unitaria tali vettori, essi risulteranno completamente coincidenti. A meno del segno dello scalare $\det(\mathbf{J})$

dunque la normale alla superficie nelle due parametrizzazioni coincide. Questo prova l'asserto.

Un'ulteriore considerazione può farsi a riguardo del denominatore dell'espressione di \mathbf{n} . Come già osservato in seguito alla definizione 3.3.2 vale la relazione $\det(\mathcal{G}) = \|\mathbf{X}_1 \wedge \mathbf{X}_2\|^2$, che ora dimostriamo.

Siano ora per comodità e per non appesantire i calcoli con le notazioni differenziali:

$$\mathbf{X}_1 = \begin{pmatrix} \frac{\partial \eta_1}{\partial \theta_1} & \frac{\partial \eta_2}{\partial \theta_1} & \frac{\partial \eta_3}{\partial \theta_1} \end{pmatrix}' = (\lambda_{11} \quad \lambda_{12} \quad \lambda_{13})'$$

$$\mathbf{X}_2 = \begin{pmatrix} \frac{\partial \eta_1}{\partial \theta_2} & \frac{\partial \eta_2}{\partial \theta_2} & \frac{\partial \eta_3}{\partial \theta_2} \end{pmatrix}' = (\lambda_{21} \quad \lambda_{22} \quad \lambda_{23})'$$

dove quindi

$$\lambda_{ij} = \frac{\partial \eta_j}{\partial \theta_i}$$

e osserviamo che, secondo queste notazioni, il tensore metrico fondamentale può essere espresso come

$$\mathcal{G} = \begin{bmatrix} \left(\frac{\partial \eta_1}{\partial \theta_1}\right)^2 + \left(\frac{\partial \eta_2}{\partial \theta_1}\right)^2 + \left(\frac{\partial \eta_3}{\partial \theta_1}\right)^2 & \frac{\partial \eta_1}{\partial \theta_1} \frac{\partial \eta_1}{\partial \theta_2} + \frac{\partial \eta_2}{\partial \theta_1} \frac{\partial \eta_2}{\partial \theta_2} + \frac{\partial \eta_3}{\partial \theta_1} \frac{\partial \eta_3}{\partial \theta_2} \\ \frac{\partial \eta_1}{\partial \theta_1} \frac{\partial \eta_1}{\partial \theta_2} + \frac{\partial \eta_2}{\partial \theta_1} \frac{\partial \eta_2}{\partial \theta_2} + \frac{\partial \eta_3}{\partial \theta_1} \frac{\partial \eta_3}{\partial \theta_2} & \left(\frac{\partial \eta_1}{\partial \theta_2}\right)^2 + \left(\frac{\partial \eta_2}{\partial \theta_2}\right)^2 + \left(\frac{\partial \eta_3}{\partial \theta_2}\right)^2 \end{bmatrix} =$$

$$= \begin{bmatrix} \lambda_{11}^2 + \lambda_{12}^2 + \lambda_{13}^2 & \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} + \lambda_{13}\lambda_{23} \\ \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} + \lambda_{13}\lambda_{23} & \lambda_{21}^2 + \lambda_{22}^2 + \lambda_{23}^2 \end{bmatrix}$$

Ricaviamo dunque $\mathbf{X}_1 \wedge \mathbf{X}_2$ tramite il prodotto esterno come da Definizione 3.3.1:

$$\begin{aligned}
\mathbf{X}_1 \wedge \mathbf{X}_2 &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \end{vmatrix} = \mathbf{i}(\lambda_{12}\lambda_{23} - \lambda_{13}\lambda_{22}) + \mathbf{j}(\lambda_{13}\lambda_{21} - \lambda_{11}\lambda_{23}) + \mathbf{k}(\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21}) = \\
&= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} (\lambda_{12}\lambda_{23} - \lambda_{13}\lambda_{22}) + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} (\lambda_{13}\lambda_{21} - \lambda_{11}\lambda_{23}) + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21}) = \\
&= \begin{pmatrix} \lambda_{12}\lambda_{23} - \lambda_{13}\lambda_{22} \\ \lambda_{13}\lambda_{21} - \lambda_{11}\lambda_{23} \\ \lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21} \end{pmatrix}
\end{aligned}$$

Il modulo al quadrato di $\mathbf{X}_1 \wedge \mathbf{X}_2$ vale

$$\|\mathbf{X}_1 \wedge \mathbf{X}_2\|^2 = (\lambda_{12}\lambda_{23} - \lambda_{13}\lambda_{22})^2 + (\lambda_{13}\lambda_{21} - \lambda_{11}\lambda_{23})^2 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})^2$$

il determinante del tensore metrico vale

$$\begin{aligned}
\det(\mathcal{G}) &= (\lambda_{11}^2 + \lambda_{12}^2 + \lambda_{13}^2)(\lambda_{21}^2 + \lambda_{22}^2 + \lambda_{23}^2) - (\lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} + \lambda_{13}\lambda_{23})^2 = \\
&= \lambda_{11}^2\lambda_{21}^2 + \lambda_{11}^2\lambda_{22}^2 + \lambda_{11}^2\lambda_{23}^2 + \lambda_{12}^2\lambda_{21}^2 + \lambda_{12}^2\lambda_{22}^2 + \lambda_{12}^2\lambda_{23}^2 + \lambda_{13}^2\lambda_{21}^2 + \lambda_{13}^2\lambda_{22}^2 + \lambda_{13}^2\lambda_{23}^2 + \\
&\quad - (\lambda_{11}^2\lambda_{21}^2 + \lambda_{12}^2\lambda_{22}^2 + \lambda_{13}^2\lambda_{23}^2 + 2\lambda_{11}\lambda_{21}\lambda_{12}\lambda_{22} + 2\lambda_{11}\lambda_{21}\lambda_{13}\lambda_{23} + 2\lambda_{12}\lambda_{22}\lambda_{13}\lambda_{23}) \\
&= \lambda_{11}^2\lambda_{22}^2 + \lambda_{11}^2\lambda_{23}^2 + \lambda_{12}^2\lambda_{21}^2 + \lambda_{12}^2\lambda_{23}^2 + \lambda_{13}^2\lambda_{22}^2 - 2\lambda_{11}\lambda_{21}\lambda_{12}\lambda_{22} - 2\lambda_{11}\lambda_{21}\lambda_{13}\lambda_{23} - 2\lambda_{12}\lambda_{22}\lambda_{13}\lambda_{23} = \\
&= (\lambda_{12}\lambda_{23} - \lambda_{13}\lambda_{22})^2 + (\lambda_{13}\lambda_{21} - \lambda_{11}\lambda_{23})^2 + (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})^2 \equiv \|\mathbf{X}_1 \wedge \mathbf{X}_2\|^2
\end{aligned}$$

questo prova l'asserto.

Un'osservazione interessante è che dunque anche $\|\mathbf{X}_1 \wedge \mathbf{X}_2\|$ è indipendente dalla scelta della parametrizzazione, in quanto funzione soltanto di un invariante del tensore metrico tramite la relazione $\|\mathbf{X}_1 \wedge \mathbf{X}_2\| = \sqrt{\det(\mathcal{G})}$.

Nota: tramite un analogo ragionamento è possibile anche dimostrare il *Theorema Egregium* di Gauss.

Abbiamo quindi dimostrato che la normale è indipendente dalla parametrizzazione.

Si osserva inoltre che la misura proposta da Bates e Watts è effettivamente coincidente con la misura di curvatura normale nota dalla Geometria Differenziale. Si rimanda all'appendice per una prova di questo fatto.

Ci riferiremo dunque a K_h^N come alla **curvatura intrinseca** nella direzione della tangente associata a \mathbf{h} .

Per quanto riguarda la componente $\ddot{\eta}_h^T$, essa può essere convertita a sua volta in una misura di **curvatura indotta dai parametri** lungo la direzione \mathbf{h} come

$$K_h^T = \frac{\|\ddot{\eta}_h^T\|}{\|\dot{\eta}_h\|^2}$$

In contrasto con la curvatura intrinseca, la curvatura indotta dai parametri dipende dalla particolare parametrizzazione scelta, e dunque una riparametrizzazione non lineare del modello da $\boldsymbol{\theta}$ ad un qualche $\boldsymbol{\beta} = g(\boldsymbol{\theta})$ può cambiare considerevolmente tale curvatura.

Effetti della replicazione di un disegno sperimentale sulla curvatura

Un disegno sperimentale (o schema di prove, piano degli esperimenti) è un insieme di combinazioni di livelli di fattori (o circostanze, condizioni, cause) che il ricercatore pone in essere nell'ordine di idee di effettuare un esperimento da cui trarre delle informazioni circa il comportamento di una variabile di interesse, che si suppone essere influenzata da detti fattori; oltre allo sperimentare diverse combinazioni dei livelli di questi fattori è possibile replicare un disegno sperimentale, ovvero ripetere due o più volte le medesime combinazioni qualora sia possibile. Variabile di interesse e fattori sono legati da un modello di regressione,

spesso non lineare, e dunque la metodologia della programmazione degli esperimenti rientra nei possibili campi di applicazione dell'oggetto di questa tesi. Si può vedere come il replicare completamente un disegno r volte ha l'effetto benefico di ridurre la curvatura in un qualunque punto ed in una qualunque direzione sul luogo delle soluzioni di un fattore pari a $1/\sqrt{r}$. Questo può essere verificato in modo semplice, dal momento che per il disegno replicato i vettori $\dot{\eta}_h^*$ e $\ddot{\eta}_h^*$ possono sempre essere scritti come r copie di $\dot{\eta}_h$ e $\ddot{\eta}_h$; per esempio $\dot{\eta}_h^* = (\dot{\eta}_h', \dot{\eta}_h', \dots, \dot{\eta}_h)'$. Segue che $\|\dot{\eta}_h^*\| = \sqrt{r} \|\dot{\eta}_h\|$ e $\|\ddot{\eta}_h^*\| = \sqrt{r} \|\ddot{\eta}_h\|$, e quindi $K_h^* = \|\ddot{\eta}_h^*\| / \|\dot{\eta}_h^*\|^2 = K_h / \sqrt{r}$. In conseguenza di questo fatto ambedue le curvature tangenziale e normale vengono ad essere ridotte del fattore $1/\sqrt{r}$. L'interpretazione statistica è che la replicazione porta ad un numero maggiore di informazioni relativamente alla variabile dipendente, con una conseguente riduzione dell'errore nella stima; per calcolare le misure di curvatura è necessario valutare le derivate in corrispondenza del punto di stima parametrica, e se questo è più preciso il luogo delle soluzioni avrà una più elevata "definizione" e ne conseguirà pertanto una più ridotta curvatura in entrambi i sensi considerati.

Diverse implicazioni della curvatura intrinseca e di quella indotta dai parametri

È importate isolare le due curvature in quanto esse influiscono in modo differente sulle regioni di confidenza della approssimazione lineare. Come già detto l'approssimazione lineare combina una *assunzione planare* ed una di *coordinate uniformi*; la prima è relativa alla sostituzione del luogo delle soluzioni con il piano tangente nel punto di stima, e tale approssimazione non sarà accettabile se la massima curvatura intrinseca è grande in $\eta(\hat{\theta})$. Similmente, la seconda implica la sostituzione sul piano tangente delle linee parametriche (curve) con una griglia di linee rette, parallele ed equispaziate, e questa approssimazione non sarà accettabile qualora la massima curvatura indotta dei parametri sia grande in $\dot{\eta}(\hat{\theta})$.

È importante puntualizzare che è necessario che l'assunzione planare sia sufficientemente soddisfatta prima che abbia senso discutere dell'accettabilità dell'assunzione di coordinate uniformi, fortunatamente, nella pratica ed in letteratura, questo è praticamente sempre verificato.

4.4 Misure relative di non linearità

Le quantità scalari K_h^N e K_h^T misurano, rispettivamente, la non linearità intrinseca del luogo delle soluzioni lungo la direzione η_h e l'apparente non linearità indotta dalla parametrizzazione del modello ma non indicano, così come sono, se tale non linearità influenzerà in modo rilevante l'inferenza basata sull'approssimazione lineare. Per poter asserire qualcosa a riguardo è necessario convertire queste due misure in *curvature relative*, che chiameremo γ_h^N e γ_h^T , che siano in un qualche senso standardizzate di modo da essere invarianti rispetto a cambiamenti di scala nella risposta. Potremo allora impiegare queste curvature relative per confrontare non solo differenti parametrizzazioni per un dato problema, ma anche differenti data set per lo stesso o diversi modelli.

Da quanto abbiamo già osservato siamo in grado di desumere che dividere risposta e modello per un fattore moltiplicherà K_h^N e K_h^T per quello stesso fattore. Un conveniente "fattore di scala" è lo **SR** introdotto in precedenza, ovvero $\rho = \sigma\sqrt{p}$, dove σ è la radice quadrata della somma dei residui al quadrato (**RSS**, ovvero Residual Sum of Squares, valore ottenibile dall'output di ogni applicativo in seguito alla stima dei parametri di un modello) divisa per $n - p$ (ovvero i gradi di libertà). Questo perché, nel caso lineare, la regione di confidenza congiunta a livello $(1 - \alpha)$ è data da quell'insieme di θ tali che

$$\|\eta(\theta) - \eta(\hat{\theta})\| \leq \rho \sqrt{F(p, n - p; \alpha)}$$

Per la dimostrazione ed ulteriori approfondimenti a riguardo dell'applicazione di questo principio nell'ambito dei modelli non lineari si rimanda all'articolo di Zanella (1975) "Sulle regioni di confidenza per i parametri dei modelli non lineari II", pagine 1 e sgg, e a Seber (1989) "Nonlinear Regression", p. 24.

Dunque la **curvatura intrinseca relativa** e la **curvatura indotta dai parametri relativa** nella direzione $\eta_{\mathbf{h}}$ risultano definite come

$$\begin{aligned}\gamma_{\mathbf{h}}^{\mathbf{N}} &= K_{\mathbf{h}}^{\mathbf{N}} \rho \\ \gamma_{\mathbf{h}}^{\mathbf{T}} &= K_{\mathbf{h}}^{\mathbf{T}} \rho\end{aligned}$$

Possiamo interpretare questa ponderazione come segue: dal momento che $K_{\mathbf{h}}^{\mathbf{N}}$ è la curvatura intrinseca, lo scalare $1/K_{\mathbf{h}}^{\mathbf{N}}$ individuerà il raggio di curvatura del luogo delle soluzioni nella direzione \mathbf{h} sul piano tangente a $\eta(\boldsymbol{\theta}_0)$; in questo modo il centro del cerchio osculatore viene a trovarsi in corrispondenza del vertice della componente $\ddot{\eta}_{\mathbf{h}}^{\mathbf{N}}$. I "raggi" $1/K_{\mathbf{h}}^{\mathbf{N}}$ e $1/K_{\mathbf{h}}^{\mathbf{T}}$ vengono confrontati con lo **SR**, così piccoli valori di $\gamma_{\mathbf{h}}^{\mathbf{N}}$ indicheranno che ρ è piccolo rispetto al raggio $1/K_{\mathbf{h}}^{\mathbf{N}}$, e grandi valori di $\gamma_{\mathbf{h}}^{\mathbf{N}}$ che ρ è grande rispetto a $1/K_{\mathbf{h}}^{\mathbf{N}}$ (lo stesso dicasi per $\gamma_{\mathbf{h}}^{\mathbf{T}}$). Per determinare l'impatto della non linearità su una particolare regione di confidenza, la curvatura così ponderata dev'essere confrontata con $1/\sqrt{F(p, n-p; \alpha)}$, che può essere visto come il raggio di curvatura della regione di confidenza al $100(1-\alpha)\%$. Se, per tutte le direzioni \mathbf{h} , $\gamma_{\mathbf{h}}^{\mathbf{N}}$ è piccolo rispetto a $1/\sqrt{F(p, n-p; \alpha)}$, allora il luogo delle soluzioni sarà relativamente piatto lungo la regione di confidenza, e quindi l'assunzione planare potrà essere accettata con ragionevole sicurezza. Similmente se $\gamma_{\mathbf{h}}^{\mathbf{T}}$ è piccolo rispetto a $1/\sqrt{F(p, n-p; \alpha)}$ allora l'assunzione di coordinate uniformi è accettabile nella regione di interesse.

Le curvature possono essere calcolate direttamente impiegando i dati già standardizzati, ovvero dove \mathbf{y} , $\boldsymbol{\eta}$, $\mathbf{e} = \mathbf{y} - \hat{\boldsymbol{\eta}}$, $\dot{\mathbf{V}}$ e $\ddot{\mathbf{V}}$ siano stati divisi per ρ . Per

convenienza d'ora in poi si assumerà che questo sia già stato fatto, e quindi d'ora in poi le misure di curvatura a cui si farà riferimento saranno quelle relative, ovvero $\gamma_{\mathbf{h}}^{\mathbf{N}}$ e $\gamma_{\mathbf{h}}^{\mathbf{T}}$.

Calcolo delle misure di curvatura relative

L'ottenimento delle misure di curvatura di non linearità che trattiamo può essere molto semplificato operando prima opportune trasformazioni lineari delle coordinate dello spazio campionario e di quello parametrico. Dal momento che siamo interessati a determinare la lunghezza delle componenti $\ddot{\eta}_{\mathbf{h}}^{\mathbf{N}}$ e $\ddot{\eta}_{\mathbf{h}}^{\mathbf{T}}$ ruotiamo le coordinate dello spazio campionario di modo che i primi p assi coordinati risultino paralleli al piano tangente definito in precedenza ed i restanti $n-p$ siano ortogonali allo stesso.

Ciò può essere ottenuto pre-moltiplicando tutti i vettori nello spazio campionario per una matrice ortogonale \mathbf{Q}' ($n \times n$), dove \mathbf{Q} è parte della *Scomposizione QR* della matrice $\dot{\mathbf{V}}$, ovvero tale per cui

$$\dot{\mathbf{V}} = \mathbf{Q}\mathbf{R} = \mathbf{Q} \begin{pmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{pmatrix}$$

e dove $\tilde{\mathbf{R}}$ è una matrice triangolare alta $p \times p$ e $\mathbf{0}$ la matrice identicamente nulla di dimensione $(n-p) \times p$.

Trasformiamo inoltre le coordinate nello spazio parametrico da $\boldsymbol{\theta}$ a

$$\boldsymbol{\Phi} = \tilde{\mathbf{R}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

di modo che le derivate $\mathbf{u}_i = \partial \eta / \partial \phi_i |_{\boldsymbol{\Phi}=0}$ siano vettori ortogonali unitari. Tali vettori compongono in effetti le prime p colonne di \mathbf{Q} in quanto

$$\begin{aligned}\dot{\mathbf{U}} &= \left. \frac{d\eta}{d\mathbf{\Phi}} \right|_{\mathbf{\Phi}=0} = \left. \frac{d\eta}{d\mathbf{\theta}} \right|_{\mathbf{\theta}=\hat{\mathbf{\theta}}} \left. \frac{d\mathbf{\theta}}{d\mathbf{\Phi}} \right|_{\mathbf{\Phi}=0} = \\ &= \mathbf{Q} \left(\begin{array}{c} \tilde{\mathbf{R}} \\ \mathbf{0} \end{array} \right) (\tilde{\mathbf{R}})^{-1} = \mathbf{Q} \left(\begin{array}{c} \mathbf{I} \\ \mathbf{0} \end{array} \right)\end{aligned}$$

e possiamo così ripartire \mathbf{Q} come $\mathbf{Q} = [\dot{\mathbf{U}} | \mathbf{N}]$. Sia ora $\mathbf{L} = \tilde{\mathbf{R}}^{-1}$ di modo che $\mathbf{\theta} = \hat{\mathbf{\theta}} + \mathbf{L}\mathbf{\Phi}$, siamo allora in grado di calcolare i vettori delle derivate seconde $\mathbf{u}_{ij} = \left. \partial^2 \eta / \partial \phi_i \partial \phi_j \right|_0$ come $\mathbf{u}_{ij} = \sum_{r,s} \mathbf{v}_{rs} l_{ri} l_{sj}$, e la matrice dei vettori delle derivate seconde nelle coordinate $\mathbf{\Phi}$ come

$$\ddot{\mathbf{U}}_t = \mathbf{L}' \ddot{\mathbf{V}}_t \mathbf{L}.$$

Nell'ultima espressione \mathbf{L} e la sua trasposta \mathbf{L}' sono matrici $p \times p$, e quindi la pre- e post-moltiplicazione dell'array $n \times p \times p$ $\ddot{\mathbf{V}}$ coinvolge la moltiplicazione di ognuna delle n $p \times p$ facce, producendo così un altro array $n \times p \times p$. Per determinare le misure che ci interessano in termini delle coordinate dello spazio campionario ruotato un passo ulteriore consiste nel pre-moltiplicare $\ddot{\mathbf{U}}$ per la matrice $(n \times n)$ \mathbf{Q}' , ottenendo così

$$\ddot{\mathbf{A}} = [\mathbf{Q}'] [\ddot{\mathbf{U}}] = [\dot{\mathbf{U}} | \mathbf{N}] [\ddot{\mathbf{U}}] = \ddot{\mathbf{A}}^T | \ddot{\mathbf{A}}^N$$

dove $\ddot{\mathbf{A}}^T$ consiste delle prime p facce di $\ddot{\mathbf{A}}$, e $\ddot{\mathbf{A}}^N$ delle restanti $n-p$. Con le parentesi quadre si sono indicate le moltiplicazioni in cui la somma è sull'indice t dei termini \mathbf{u}_{ij} .

Il vantaggio dell'impiego dei parametri $\mathbf{\Phi}$ è che se $\mathbf{d} = (d_1, d_2, \dots, d_p)'$ è un vettore unitario, allora la linea corrispondente $\dot{\eta}_{Ld}$ possiede $\|\dot{\eta}_{Ld}\| = 1$, quindi

$$\gamma_{Ld}^T = \left\| (\mathbf{d}'\ddot{\mathbf{U}}\mathbf{d})^T \right\| = \left\| \mathbf{d}'\ddot{\mathbf{A}}^T\mathbf{d} \right\|$$

$$\gamma_{Ld}^N = \left\| (\mathbf{d}'\ddot{\mathbf{U}}\mathbf{d})^N \right\| = \left\| \mathbf{d}'\ddot{\mathbf{A}}^N\mathbf{d} \right\|$$

Analizzando queste espressioni ci si rende conto che $\gamma_{\mathbf{h}}^T$ e $\gamma_{\mathbf{h}}^N$ risultano essere indipendenti dalla lunghezza di \mathbf{h} (o $L\mathbf{d}$). Dal momento che la trasformazione da $\boldsymbol{\theta}$ a $\boldsymbol{\Phi}$ è di tipo lineare, le *lifted lines* saranno le medesime per entrambe le parametrizzazioni, e dunque ci basterà considerare soltanto vettori unitari \mathbf{d} per ottenere le curvature relative in tutte le possibili direzioni.

Il vantaggio delle \mathbf{Q} -coordinate è che lo spazio campionario è ridefinito con rispetto al piano tangente, caratterizzato da $\dot{\mathbf{U}}$, e al suo complemento normale \mathbf{N} . Ogni vettore di accelerazione \mathbf{a}_{ij} può quindi essere scritto come la somma delle componenti ortogonali \mathbf{a}_{ij}^T e \mathbf{a}_{ij}^N (queste ultime indipendenti da qualunque effetto indotto dai parametri). Invece i vettori \mathbf{a}_{ij}^T che formano l'array $\ddot{\mathbf{A}}^T$ sono influenzati dalla parametrizzazione, e quindi $\ddot{\mathbf{A}}^T$ raccoglie e quantifica questo effetto. Una perfetta riparametrizzazione dovrebbe quindi essere quella che riduce ogni elemento di questo array a zero.

Interpretazione geometrica delle componenti dell'array $\ddot{\mathbf{A}}^T$

Nella sua tesi di dottorato D.M.Bates ha dimostrato che ognuno dei termini a_{ijk} nell'array $\ddot{\mathbf{A}}^T$ ha una diretta interpretazione geometrica riguardante il comportamento delle proiezioni delle linee parametriche sul piano tangente approssimante. I termini della forma a_{iii} rappresentano “non-uniformità di scala” dovuta al parametro ϕ_i , i termini della forma $a_{iji} = a_{ijj}$ rappresentano ciò che Bates chiama *fanning*, e che potrebbe essere tradotto in italiano come “disposizione a ventaglio”, delle linee parametriche relative a ϕ_j nella direzione ϕ_i ; i termini della forma a_{ijj} rappresentano un *arching*, traducibile come flessione ad arco, delle linee

parametriche relative a ϕ_j nella direzione ϕ_i , ed infine i restanti termini della forma generica a_{ijk} (con $i \neq j \neq k$) rappresentano la torsione della superficie parametrica (ϕ_j, ϕ_k) . Dunque l'approccio geometrico allo studio di $\ddot{\mathbf{A}}^T$ può fornire una visuale più approfondita nelle problematiche legate alla stima non lineare.

Un algoritmo per la determinazione delle misure relative di curvatura

Siamo ora in possesso degli strumenti per poter calcolare le curvature **IN** e **PE** relative lungo una particolare direzione dello spazio parametrico \mathbf{h} . Da queste possiamo ottenere misure del grado di non linearità calcolando la **massima curvatura intrinseca** Γ^N e la **massima curvatura indotta dai parametri** Γ^T . Bates e Watts nel loro articolo affermano di non avere a disposizione formule per calcolare in modo diretto Γ^N e Γ^T , tranne che in casi particolari, ma presentano un semplice algoritmo iterativo per l'ottenimento di queste quantità.

Nota: potrebbe essere interessante indagare sull'esistenza di una relazione esplicita che leghi tali massime curvature agli autovalori dell'operatore di Weingarten (Teorema 3.4.1) che individuano le curvature principali; nella letteratura esaminata per la stesura di questa tesi non è stato trovato riferimento a questa eventualità.

Dal momento che stiamo considerando soltanto vettori di norma unitaria \mathbf{d} , un particolare vettore \mathbf{d}^N individuerà un massimo locale di $\gamma_{\mathbf{d}}^N$ qualora il gradiente $\nabla \gamma_{\mathbf{d}}^N$ abbia la stessa direzione di \mathbf{d}^N . Se il valore in esame non rappresenta un massimo allora si passa al tentativo seguente impostando come valore la direzione del gradiente. Per comodità gli autori utilizzano $\nabla (\gamma_{\mathbf{d}}^N)^2$ anziché $\nabla \gamma_{\mathbf{d}}^N$, il che non comporta problemi dal momento che questi due gradienti hanno la stessa direzione. Formalmente l'algoritmo può essere espresso come segue:

- (1) scelta di una direzione iniziale \mathbf{d}_i
- (2) calcolo delle quantità $\mathbf{g}_i = \nabla(\gamma^N)^2 \Big|_{\mathbf{d}_i} = 4[\mathbf{d}'_i \ddot{\mathbf{A}}^N \mathbf{d}_i]' [\ddot{\mathbf{A}}^N \mathbf{d}_i]$ e $\tilde{\mathbf{g}}_i = \mathbf{g}_i / \|\mathbf{g}_i\|$
- (3) se $\tilde{\mathbf{g}}'_i \mathbf{d}_i < 1 - \zeta$ allora $\mathbf{d}_{i+1} = \tilde{\mathbf{g}}_i$ e si riprenda da (2), altrimenti $\Gamma^N = \|\mathbf{d}'_i \ddot{\mathbf{A}}^N \mathbf{d}_i\|$.

Il criterio di convergenza ζ nel passo (3) indica quanto vicini nella direzione richiediamo che siano \mathbf{d}_i ed il gradiente. Dal momento che sia \mathbf{d}_i che $\tilde{\mathbf{g}}_i$ sono vettori di norma unitaria il loro prodotto interno sarà 1 se essi sono esattamente nella stessa direzione; un valore di $\zeta = 0.0001$, ad esempio, imporrà che tale prodotto debba risultare maggiore di 0.9999 perché la convergenza sia raggiunta.

Gli autori apportano nello stesso articolo in cui lo presentano una correzione all'algoritmo in virtù dell'osservazione di una accentuata oscillazione intorno all'ottimo man mano che ci si avvicina alla convergenza. Tale correzione consiste nella modifica del punto (3) sostituendo a $\mathbf{d}_{i+1} = \tilde{\mathbf{g}}_i$ l'espressione $\mathbf{d}_{i+1} = (3\tilde{\mathbf{g}}_i + \mathbf{d}_i) / \|3\tilde{\mathbf{g}}_i + \mathbf{d}_i\|$; essi osservano inoltre che impiegando la Scomposizione QR con una tecnica di pivoting per calcolare $\ddot{\mathbf{A}}$ l'algoritmo tendeva ad individuare un \mathbf{d}^N vicino a $(0,0,\dots,0,1)'$, fatto questo che li porta ad impostare la direzione iniziale \mathbf{d}_i pari a $\mathbf{d}_0 = (0,0,\dots,0,1)'$. La massima curvatura indotta dai parametri Γ^T viene calcolata dallo stesso algoritmo sostituendo $\ddot{\mathbf{A}}^T$ a $\ddot{\mathbf{A}}^N$ nei passi (2) e (3).

Gli autori affermano che con questo valore di partenza ed il criterio di $\zeta = 0.0001$ l'algoritmo modificato ha raggiunto la convergenza solitamente in quattro iterazioni, e non ne ha mai richieste più di dieci per alcuno degli esempi che portano nell'articolo. Per assicurarsi inoltre di avere raggiunto un massimo globale, e non soltanto locale, si sono serviti di una tecnica che prevede il calcolo di *bounds* di Γ descritta nell'appendice dell'articolo, se il massimo fosse caduto all'interno di tale regione allora sarebbe stato accettato come massimo globale.

Nelle simulazioni che seguiranno le misure di curvatura relative sono state calcolate tramite la funzione *rms.curv*, presente nel package MASS 7.2-30 in ambiente R 2.4.1, basata su leggera modificazione di questo algoritmo, che

contempla il calcolo di una media potenziata di ordine 2 delle curvature in ogni direzione come misura di riferimento; il risultato di questa operazione è l'ottenimento di misure "medie" delle curvature **IN** e **PE** nell'intorno del punto di stima parametrica, nell'ottica di una valutazione che non consideri solamente la massima curvatura possibile, la quale potrebbe presentarsi particolarmente elevata come caso particolare unicamente in una determinata direzione, mentre lungo le restanti potrebbe risultare contenuta.

Calcolo della misura di distorsione di Box

La formula derivata da Box nel 1971 per il calcolo della distorsione nelle stime dei parametri per un modello non lineare con risposta univariata è la seguente

$$Bias(\hat{\boldsymbol{\theta}}) = E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{-\sigma^2}{2} \left(\sum_{u=1}^n \mathbf{F}_u \mathbf{F}_u' \right)^{-1} \sum_{t=1}^n \mathbf{F}_t tr \left[\left(\sum_{v=1}^n \mathbf{F}_v \mathbf{F}_v' \right)^{-1} \mathbf{H}_t \right]$$

dove \mathbf{F}_u è ρ volte la u -esima riga di $\dot{\mathbf{V}}$ e \mathbf{H}_t è ρ volte la t -esima faccia dell'array $\ddot{\mathbf{V}}$, calcolati in corrispondenza delle condizioni sperimentali. All'atto pratico l'utilizzo della formula implica la sostituzione di $\hat{\sigma}^2$ e $\hat{\boldsymbol{\theta}}$ alle quantità ignote.

La misura così risultante è un vettore $p \times 1$ che rappresenta una stima dello scostamento delle stime dei parametri dal vero valore parametrico.

Bates e Watts hanno mostrato che la distorsione della stima secondo i LS del parametro j -esimo (ovvero l'elemento $\hat{\theta}_j$ di $\hat{\boldsymbol{\theta}}$) è semplicemente una trasformata dell'elemento diagonale medio di una componente dell'array delle derivate seconde; formalmente hanno dimostrato che

$$E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{-\sigma^2}{2\hat{\sigma}^2 p} \mathbf{L} \left(\sum_{i=1}^p \mathbf{a}_{ii}^T \right)$$

e analogamente, nei parametri $\boldsymbol{\Phi}$

$$E(\hat{\Phi} - \Phi) = \frac{-\sigma^2}{2\hat{\sigma}^2 p} \sum_{i=1}^p \mathbf{a}_{ii}^T$$

quindi, ribadendo il concetto, la distorsione relativa al parametro ϕ_j altro non è che una ponderazione dell'elemento diagonale medio della j -esima faccia di $\ddot{\mathbf{A}}^T$. Questo pone l'accento sul fatto che, per quanto riguarda la formula di Box, la distorsione è una proprietà strettamente legata alla parametrizzazione, e può essere ridotta o eliminata tramite una conveniente riparametrizzazione del modello, come già accennato in precedenza. Sostituendo $\hat{\sigma}^2$ a σ^2 otteniamo la stima del bias.

In letteratura è particolarmente impiegato il *percentage bias* basato su questa stima, che si può ricavare rapportando elemento per elemento il vettore $Bias(\hat{\theta})$ al vettore $\hat{\theta}$ dei parametri stimati, moltiplicando il risultato per 100. Le componenti in valore assoluto del vettore risultante forniscono una misura percentuale della distorsione di ogni singolo parametro, e una regola operativa suggerisce come 1% la misura oltre la quale la distorsione indica un comportamento eccessivamente non lineare del parametro stesso.

È possibile ottenere la misura della distorsione dei nuovi parametri ottenuti da una riparametrizzazione degli originari come segue: si consideri una riparametrizzazione dai parametri θ ai parametri δ tramite una funzione regolare del tipo $\delta = g(\theta)$. Allora si può dimostrare che

$$Bias(\hat{\delta}) = \mathbf{g}' Bias(\hat{\theta}) + \frac{1}{2} tr[\mathbf{M}Cov(\hat{\theta})]$$

dove \mathbf{g} è il vettore $p \times 1$ delle derivate prime e \mathbf{M} la matrice $p \times p$ delle derivate seconde di $g(\theta)$ rispetto a θ , valutate in $\hat{\theta}$. $Cov(\hat{\theta})$ è la matrice asintotica di varianze e covarianze di $\hat{\theta}$, ottenibile dall'inversa della matrice di informazione; tale matrice, sotto le ipotesi in cui stiamo lavorando, possiede componenti calcolabili tramite l'espressione

$${}_{\theta} \mathbf{I}_{ij} = \frac{1}{\sigma^2} \left[\sum_{t=1}^n \frac{\partial \eta(\mathbf{x}_t, \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \eta(\mathbf{x}_t, \boldsymbol{\theta})}{\partial \theta_j} \right]$$

Si noti come la matrice di informazione del Fisher risulti coincidente con il tensore metrico \mathcal{G} della varietà indotta dal modello non lineare a meno della costante moltiplicativa $1/\sigma^2$. Potrebbe costituire un interessante campo di indagine il legame tra la matrice di varianze e covarianze dei parametri e le misure di curvatura note dalla Geometria Differenziale e presentate nelle loro espressioni, funzionali del tensore metrico, nel paragrafo 3.4.

Ulteriori considerazioni sulle implicazioni della non linearità intrinseca

La componente di non linearità **IN** è informativa della curvatura del luogo delle soluzioni, come già detto, e risulta definita in corrispondenza della specificazione di una combinazione modello-data set. Nella la quasi totalità degli esempi presenti in letteratura l'effetto di tale curvatura risulta trascurabile, ma è necessario portare alcune ulteriori considerazioni al riguardo. La portata dell'impatto della non linearità intrinseca risulta particolarmente evidente nel momento in cui il ricercatore decida di impiegare il modello stimato per la previsione dei valori della variabile dipendente e nella determinazione di limiti di confidenza per gli stessi. A tal fine, delle due misure di non linearità introdotte da Bates e Watts, solamente la curvatura **IN** risulta rilevante, come mostreremo.

Si consideri la seguente espressione per la previsione dei valori \hat{y} relativa ad un modello di regressione non lineare:

$$\hat{y} = f(\mathbf{X}, \hat{\boldsymbol{\theta}})$$

dove i simboli hanno i significati noti. È importante sottolineare come la previsione \hat{y} per un dato insieme di condizioni sperimentali \mathbf{X} risulterà inalterata in seguito ad

una qualsivoglia riparametrizzazione $\delta = g(\theta)$, tale quindi per cui il modello stimato possa essere riscritto come

$$\hat{y} = f(\mathbf{X}, g^{-1}(\hat{\delta}))$$

in quanto la relazione di riparametrizzazione è funzione soltanto dei parametri originari, e non delle condizioni sperimentali.

Dal momento che i valori \hat{y} sono indipendenti dalla parametrizzazione scelta, segue che la distorsione della stima degli stessi non potrà essere legata alla non linearità indotta dai parametri la quale, ricordiamo, potrà sempre, almeno teoricamente, essere ridotta a piacere da una conveniente riparametrizzazione. Quindi il bias nella previsione risulta essere collegato soltanto alla componente di non linearità intrinseca. Box deriva una espressione per il calcolo di questa distorsione facendo uso di una espansione in serie di Taylor troncata al secondo ordine

$$Bias(\hat{y}) = \mathbf{F}'Bias(\hat{\theta}) + \frac{1}{2}tr[\mathbf{H}Cov(\hat{\theta})]$$

dove i simboli hanno i significati già presentati. La formula assomiglia molto alla relazione fra la distorsione di due parametrizzazioni diverse, con la differenza però che le derivate sono calcolate relativamente al modello anziché alla funzione riparametrizzante. Box derivò anche una formula per la varianza dei valori previsti \hat{y} , sebbene ad un minor ordine di accuratezza della stima della distorsione:

$$Var(\hat{y}) = \mathbf{F}'Cov(\hat{\theta})\mathbf{F}$$

In Ratkowsky (1983) sono presenti numerosi esempi numerici che mostrano l'alta correlazione fra non linearità intrinseca e distorsione delle stime \hat{y} . Più grande risulta essere la curvatura del luogo delle soluzioni, maggiore sarà la distorsione

nelle stime. Questo fatto dimostra che se lo scopo principale del ricercatore è quello di impiegare il modello stimato come strumento previsivo, potrà non essere sufficiente avere una curvatura intrinseca che non ecceda la soglia critica di $1/\sqrt{F(p, n-p; \alpha)}$, ma sarà auspicabile che si trovi più al di sotto possibile di tale livello di accettabilità. Una utile indicazione per poter contenere la non linearità intrinseca è, come si è già visto, la replicazione del piano sperimentale, operazione che porta ad una riduzione di un fattore $1/\sqrt{r}$ in termini di curvatura **IN**, r essendo il numero di replicazioni del disegno.

Il t -value e le misure di curvatura di Bates e Watts

La statistica che esprime il rapporto fra la stima di un parametro secondo il criterio dei LS ed il suo standard error è nota come t -value, e viene spesso impiegata nell'esame del comportamento del modello relativamente alla significatività dei parametri. Tali statistiche vengono confrontate con una distribuzione t di Student con $n-p$ gradi di libertà e una conseguente regione di confidenza al livello $100(1-\alpha)\%$. Il criterio afferma che un t -value sufficientemente alto indica che il parametro è ben definito (ovvero è da ritenersi diverso da zero al livello di confidenza prefissato); ciò però non necessariamente significa che la stima del parametro possieda altre desiderabili proprietà statistiche, quali ad esempio un comportamento "prossimo al lineare" (distribuzione normale, minima varianza, non distorsione, e così via). Inoltre, considerando il caso dei modelli multiparametrici, il t -value può risultare basso a causa dell'alta correlazione tra i parametri stessi; nonostante in molte situazioni il t -value possa essere un utile indicatore di comportamento non lineare delle stime, la sua validità dipende da molte delle proprietà statistiche del modello, inclusa la vicinanza alla linearità nei termini presentati in questa tesi. Questo porta a suggerire che la considerazione congiunta di t -value, **IN** e **PE** sia da preferirsi al solo fare riferimento a questa statistica.

Considerazioni sulla correlazione fra le stime LS in un modello non lineare

La matrice di correlazione delle stime parametriche sarà formata da una diagonale di 1 (correlazione di un parametro con sé stesso) e di elementi fuori dalla diagonale principale che si auspica siano il più possibili prossimi allo zero. Come già detto per quanto riguardava il t -value, non c'è ragione di credere che una alta correlazione fra i parametri sia indicativa di comportamento non lineare (distorsione, eccesso di varianza, distribuzione non normale) delle stime dei parametri in un modello non lineare. È semplice verificare come, anche per un modello strettamente lineare, i coefficienti di correlazione possano essere resi grandi a piacere semplicemente tramite una traslazione delle variabili indipendenti (ovvero l'aggiunta di una costante additiva a tutti i regressori). Quindi, per il modello non lineare, un'alta correlazione fra le stime secondo i LS non può rappresentare una indicazione significativa di comportamento non lineare delle stesse. La diffusa convinzione del contrario trova forse le sue motivazioni nel fatto che modelli non lineari con un "cattivo" comportamento (ovvero grande lontananza dalla linearità) spesso presentano correlazioni molto elevate nelle stime.

Benefici della riparametrizzazione

A fronte di una curvatura intrinseca sufficientemente ridotta, e se il ricercatore ha successo nell'individuare una opportuna riparametrizzazione che riconduca ad un livello accettabile la curvatura indotta dai parametri, si avrà dunque a disposizione un modello per il quale le stime parametriche si comporteranno in modo soddisfacentemente prossimo al lineare. I benefici principali sono riassumibili come segue:

1. Le stime LS potranno essere ottenute agevolmente ed in un numero ridotto di passaggi da un algoritmo di stima come quelli già descritti, ricordiamo infatti che per il modello lineare tali procedure necessitano di una singola operazione per

giungere alla convergenza. Quindi maggiore la vicinanza alla linearità, minore il numero di iterazioni necessarie, anche partendo da valori iniziali (necessari per questi algoritmi) che non siano particolarmente precisi. Un modello non lineare molto lontano dalla linearità potrebbe portare anche alla non convergenza dell'algoritmo, a meno che non sia possibile fornire valori iniziali molto prossimi a quelli di convergenza.

2. Dal momento che gli stimatori LS saranno quasi non distorti e quasi normalmente distribuiti, e siccome le varianze delle stime saranno prossime al limite inferiore (fornito, lo ricordiamo, dalla matrice asintotica di varianze e covarianze), questo consentirà di poter impiegare con relativa sicurezza semplici preliminari analisi come ad esempio il confronto tramite test t delle stime parametriche tra data-set differenti. Si ricordi ad ogni modo che queste considerazioni potranno essere solo preliminari, per i motivi già discussi in precedenza.
3. Quanto più un modello non lineare si avvicina ad un comportamento lineare, tanto più validi saranno in generale test statistici e procedure il cui uso è derivato in analogia ai modelli lineari e che assumono ipotesi di normalità delle stime. Per queste metodologie il ricercatore potrà confidare del fatto che i livelli di significatività ipotizzati siano sufficientemente ben approssimati quando il comportamento del modello è prossimo al lineare, ovvero quanto più lineare sarà il comportamento, tanto più le regioni di confidenza dei parametri assomiglieranno ad ellissoidi (ellissi in due dimensioni, iper-ellissoidi in più di tre).
4. In ultimo, l'effettivo, ma ignoto, livello di significatività risulterà più concordante con il livello nominale α al decrescere della non linearità intrinseca, dal momento che la distorsione nella somma dei residui al quadrato dipende dalla grandezza della curvatura **IN**.

4.5 Studio tramite simulazioni del comportamento delle stime secondo i LS

Le simulazioni che seguiranno sono state interamente programmate e svolte in ambiente *R* 2.4.1, l'ultima versione del software open source *R* della “*The R Foundation for Statistical Computing*”, nato come versione didattica del più noto (e commerciale) applicativo *S+*. Il software, di cui sono presenti versioni per differenti piattaforme (Windows, Macintosh, Linux, e così via), è disponibile per il download gratuito all'indirizzo <http://cran.r-project.org/>. Il grande pregio di questo applicativo è senz'altro la sua natura open source, il che consente a qualunque utilizzatore nel mondo di mettere a disposizione (e fruire gratuitamente di) package aggiuntivi creati ad hoc per particolari problemi, inoltre la trasposizione di pacchetti professionali creati per *S+* risulta particolarmente agevole; ulteriore pregio la limpidezza del linguaggio di programmazione, che presenta inoltre molti punti di contatto con altri linguaggi object-oriented, come MatLab ad esempio, e risulta di semplice apprendimento. Le enormi potenzialità di questo software e l'instancabile lavoro del suo team di sviluppo lo rendono uno strumento estremamente flessibile ed in continua via di miglioramento, per quanto, purtroppo, ancora poco utilizzato in Italia.

In quanto segue verranno analizzate le proprietà statistiche delle stime parametriche per alcune combinazioni di modelli-data set tratte dalla letteratura e rielaborate alla luce delle tecniche presentate in questa tesi; si presenteranno inoltre simulazioni concernenti la distribuzione delle stime e considerazioni sulla distorsione ed efficienza delle stesse. Si valuterà l'impatto sulle proprietà sopracitate di eventuali riparametrazioni, la cui convenienza sia stata suggerita dagli indicatori di non linearità e distorsione, dell'aumento della dimensione campionaria e della replicazione del disegno sperimentale. Le condizioni a riguardo degli errori ε_t sono le usuali (*iid* normali, di media nulla e varianza finita σ^2).

Un semplice modello non lineare monoparametrico

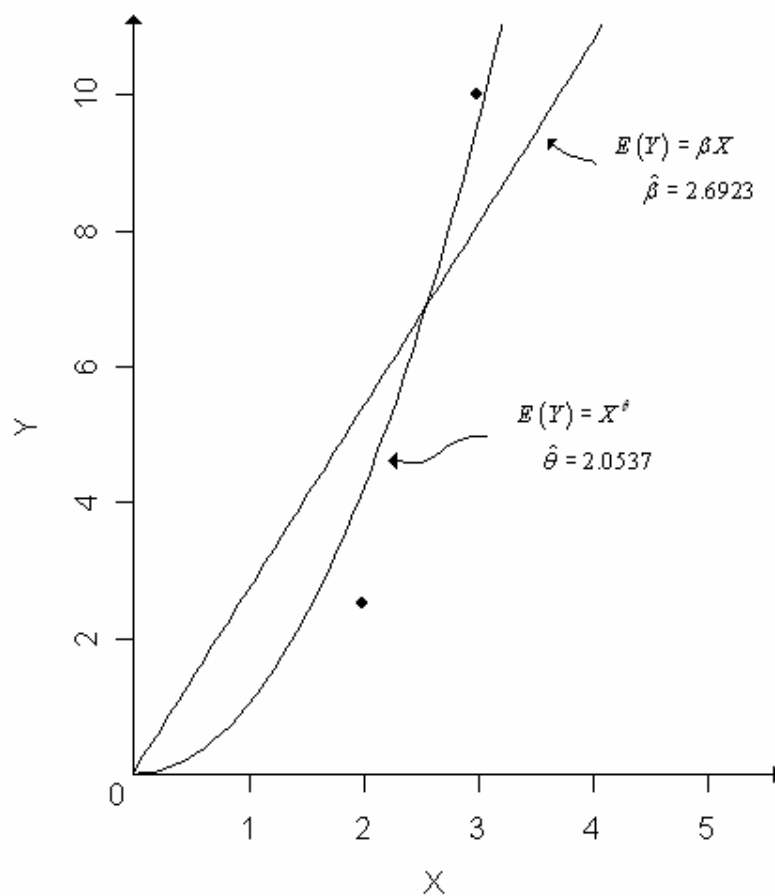
La combinazione modello-data set che segue è tratta da Ratkowsky (1983).

Modello: $Y_i = X_i^\theta + \varepsilon_i \Rightarrow \eta(\theta) = E(Y) = X^\theta$

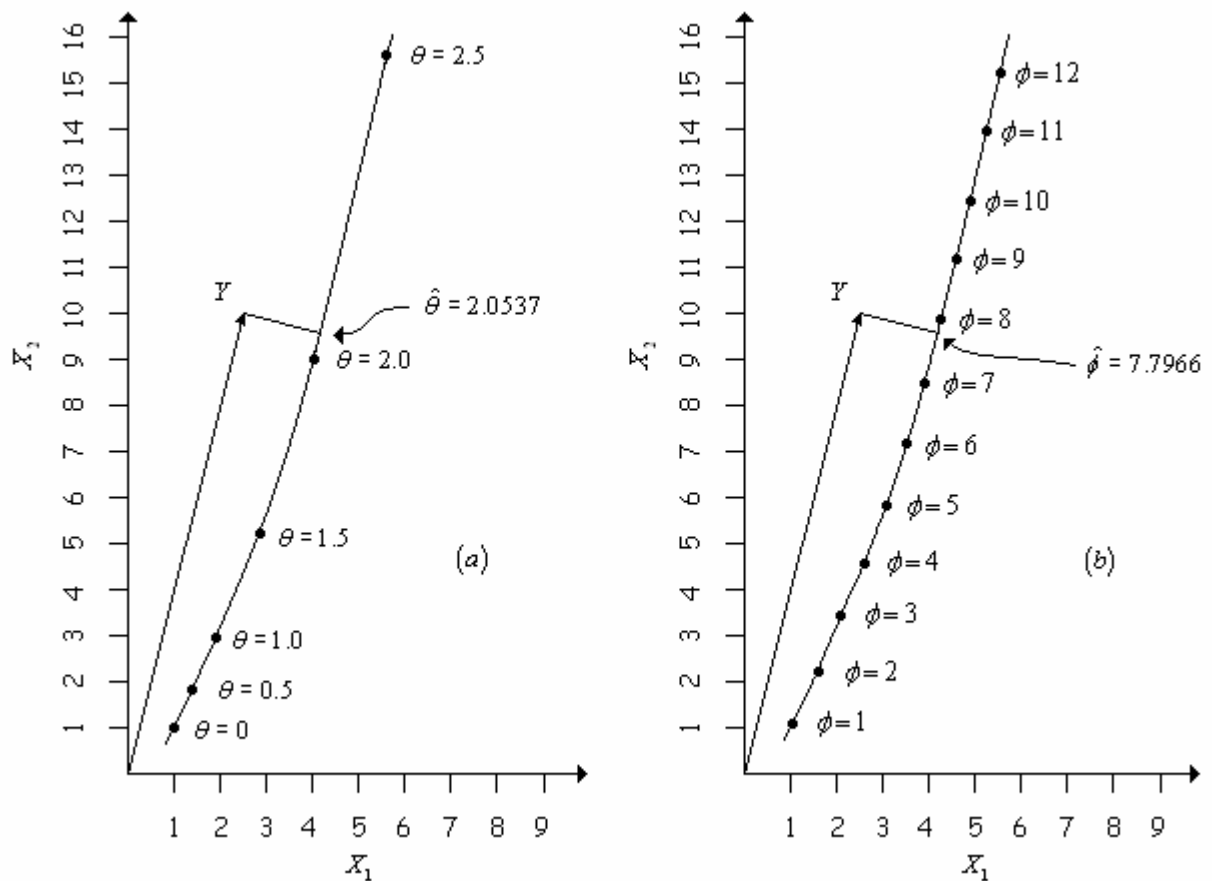
Dati:

X_i	Y_i
2	2.5
3	10

La stima secondo i LS fornisce un valore del parametro pari a $\hat{\theta} = 2.053690$, e la curva risultante è rappresentata nel seguente grafico, confrontato con un modello retta incompleto.



Anziché rappresentare graficamente i dati nel modo tradizionale si può impiegare una metodologia alternativa, che bene servirà lo scopo di interpretare, almeno qualitativamente, le considerazioni sulla non linearità presentate in questo capitolo. È possibile disegnare (figura (a)) il luogo delle soluzioni del modello, quest'ultimo essendo il luogo dei punti aventi la forma (X_1^θ, X_2^θ) , nel caso in esame $(2^\theta, 3^\theta)$, rappresentante quindi tutti i possibili valori del parametro o soluzioni del modello; per fare ciò necessitiamo dunque di un numero di assi coordinati pari alla dimensione campionaria, in questo caso due (risulta lampante come non sarà possibile sfruttare questo tipo di rappresentazione per campioni composti da più di tre elementi).



La stima LS $\hat{\theta}$ è quel punto sul luogo delle soluzioni più vicino ad Y , rappresentato dal vettore per l'origine $(2.5, 10)$. Geometricamente la si può ottenere individuando quel punto sul luogo delle soluzioni tale per cui la lunghezza della perpendicolare

dal vertice del vettore alla linea tangente alla curva in quel punto sia minima (il principio è analogo a quello per la determinazione della stima nel modello lineare). La forma del luogo delle soluzioni e la spaziatura dei valori relativi a variazioni costanti $\Delta\theta$ del parametro nelle vicinanze di $\hat{\theta}$ possono essere impiegate come misure del grado in cui il modello non lineare si discosta da quello lineare. Questa considerazione discende dal fatto che, in un modello lineare, il luogo delle soluzioni è lineare (come dimostrato nel paragrafo 1.3) e quei punti corrispondenti ad eguali incrementi $\Delta\theta$ sono essi stessi equispaziati.

Facendo riferimento ancora alla figura (a) il luogo delle soluzioni mostra essere leggermente incurvato nelle prossimità di $\hat{\theta}$, e la quantificazione di questo fatto può essere attuata tramite la misura di curvatura relativa **IN**; ci attenderemo, in conformità all'evidenza grafica, un basso grado di curvatura intrinseca. L'altro aspetto per cui un modello non lineare differisce da quello lineare coinvolge la posizione sul luogo delle soluzioni di valori di θ posti ad uguali incrementi $\Delta\theta$: nella figura (a) sono stati riportati i valori di θ corrispondenti ad incrementi di 0.5, ed è immediato vedere come essi non siano equispaziati, ma tendano ad allontanarsi sempre di più fra di loro al crescere di θ . Lo strumento adatto per quantificare questo aspetto di non linearità è la misura di curvatura relativa **PE**, poiché è la componente che risente del modo in cui i parametri appaiono nel modello, ovvero è relativa alla parametrizzazione scelta; la curvatura **PE** è una misura che si basa sulla spaziatura di $\Delta\theta$ quando θ è proiettato sul piano tangente (in questo caso sulla tangente) al luogo delle soluzioni in $\hat{\theta}$ e, come già detto in precedenza, può essere alterata anche considerevolmente tramite una opportuna riparametrizzazione.

Se ad esempio riparametriamo il modello come

$$\eta(\phi) = E(Y) = X^{\log(\phi)}$$

allora la stima secondo i LS di ϕ vale $\hat{\phi} = \exp(\hat{\theta}) = 7.7966$. La figura (b) mostra l'effetto di tale riparametrizzazione: il luogo delle soluzioni rimane inalterato, come ci si attendeva, ma ora a variazioni costanti $\Delta\phi = 1$ del parametro ϕ sul luogo delle

soluzioni si può notare come la spaziatura di questi incrementi uguali sia quasi identica lungo tutta la curva. Dunque potremmo aspettarci che il nuovo modello, che altro non è se non una riparametrizzazione di quello originario, mostri un comportamento più prossimo alla linearità rispetto al primo, come in effetti risulta essere e come verificheremo fra breve.

Procediamo dunque con il calcolo delle misure di curvatura relative di Bates e Watts per il primo modello; seguiamo la procedura descritta nel precedente paragrafo 4.4 e deriviamo le quantità necessarie allo scopo.

In questo caso $p=1$, $n=2$, $n-p=1$. Per calcolare lo **SR** $\rho = \hat{\sigma} \sqrt{p}$ è necessario conoscere la varianza residua $\hat{\sigma}^2$, ottenibile dall'output del applicativo impiegato per la stima del modello, oppure come $\left[(2.5 - 4.1517)^2 + (10 - 9.5468)^2 \right] / 1 = 2.9334$; quindi $\rho = 1.7127$. Le matrici di derivate parziali prime e seconde, standardizzate per lo **SR**, sono rispettivamente

$$\dot{\mathbf{V}} = \frac{1}{1.7127} \begin{pmatrix} (\log 2) 2^{2.0537} \\ (\log 3) 3^{2.0537} \end{pmatrix} = \begin{pmatrix} 1.6802 \\ 6.1238 \end{pmatrix}$$

$$\ddot{\mathbf{V}} = \frac{1}{1.7127} \begin{pmatrix} (\log 2)^2 2^{2.0537} \\ (\log 3)^2 3^{2.0537} \end{pmatrix} = \begin{pmatrix} 1.1646 \\ 6.7277 \end{pmatrix}$$

La matrice $\dot{\mathbf{V}}$ viene scomposta per fornire

$$\dot{\mathbf{V}} = \mathbf{QR} = \mathbf{Q} \begin{pmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{pmatrix} =$$

$$= \begin{pmatrix} -0.26460 & -0.96438 \\ -0.96438 & 0.26460 \end{pmatrix} \begin{pmatrix} -6.3500 \\ 0 \end{pmatrix}$$

L'inversa di $\tilde{\mathbf{R}}$ è $\mathbf{L} = \tilde{\mathbf{R}}^{-1} = (-6.3500)^{-1} = -0.15748$, da cui otteniamo

$$\begin{aligned}\ddot{\mathbf{U}} &= \mathbf{L}'\ddot{\mathbf{V}}\mathbf{L} = -0.15748 \begin{pmatrix} 1.1646 \\ 6.7277 \end{pmatrix} (-0.15748) = \\ &= \begin{pmatrix} 0.028882 \\ 0.166850 \end{pmatrix}\end{aligned}$$

Quindi l'array di accelerazione $\ddot{\mathbf{A}}$ è

$$\begin{aligned}\ddot{\mathbf{A}} &= \mathbf{Q}'\ddot{\mathbf{U}} = \begin{pmatrix} -0.26460 & -0.96438 \\ -0.96438 & 0.26460 \end{pmatrix} \begin{pmatrix} 0.028882 \\ 0.166850 \end{pmatrix} = \\ &= \begin{pmatrix} -0.1685 \\ 0.0163 \end{pmatrix} \left. \begin{array}{l} \} \text{le prime } p \text{ righe legate alla PE} \\ \} \text{le ultime } n - p \text{ righe relative alla IN} \end{array} \right\}\end{aligned}$$

Per un array di accelerazione comprendente un singolo elemento (caso di modelli monoparametrici) la massima curvatura è semplicemente il valore assoluto dell'elemento. Dunque otteniamo i seguenti valori di curvatura

$$\begin{aligned}\text{IN} &= 0.0163 \\ \text{PE} &= 0.1685\end{aligned}$$

che andranno confrontati con il valore critico $1/\sqrt{F}$, con $F = F(1,1;0.05) = 161.4476$, ovvero $1/\sqrt{F} = 0.0787$. La non linearità intrinseca del luogo delle soluzioni nel punto di stima parametrica è dunque accettabile secondo questo criterio, dal momento che $\text{IN} < 0.0787$ l'assunzione planare è soddisfatta, mentre poiché $\text{PE} > 0.0787$ siamo costretti a rifiutare l'assunzione di coordinate uniformi. Quest'ultima affermazione risulta in linea con i ragionamenti relativi alla spaziatura sul luogo delle soluzioni a fronte di incrementi costanti del parametro.

Per il modello riparametrizzato tramite la relazione $\phi = \exp(\theta)$ applicando la medesima procedura si ottengono i seguenti valori di curvatura:

$$\begin{aligned}\text{IN} &= 0.0163 \\ \text{PE} &= 0.0111\end{aligned}$$

La non linearità **IN** è rimasta identica, come era lecito attendersi dal momento che la riparametrizzazione non altera la forma del luogo delle soluzioni, ma la non linearità **PE** è risultata di molto ridotta, rientrando nel limite fornito dal valore critico: ora infatti $PE < 0.0787$, e quindi l'assunzione di coordinate uniformi è accettabile.

Procediamo ora al calcolo della distorsione della stima parametrica secondo la formula di Box presentata nel paragrafo 4.4, e valutiamo l'impatto della riparametrizzazione in questo senso. Dal momento che $p = 1$ molte delle operazioni matriciali della formula

$$Bias(\hat{\boldsymbol{\theta}}) = E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{-\sigma^2}{2} \left(\sum_{u=1}^n \mathbf{F}_u \mathbf{F}'_u \right)^{-1} \sum_{t=1}^n \mathbf{F}_t tr \left[\left(\sum_{v=1}^n \mathbf{F}_v \mathbf{F}'_v \right)^{-1} \mathbf{H}_t \right]$$

dove i simboli hanno i significati già presentati, vengono ad essere semplici operazioni fra scalari, semplificando così il calcolo; valori $\boldsymbol{\theta}$ e σ^2 nell'espressione devono essere sostituiti con $\hat{\theta} = 2.0537$ e $\hat{\sigma}^2 = 2.9334$ per ottenere la stima del bias, calcoliamo inoltre la matrice di varianze e covarianze della stima (necessaria per valutare la distorsione a seguito della riparametrizzazione), che in questo caso ad monoparametrico risulta coincidente con la sola varianza, tramite l'inversa dell'informazione

$${}_{\boldsymbol{\theta}} \mathbf{I}_{ij} = \frac{1}{\sigma^2} \left[\sum_{t=1}^n \frac{\partial \eta(\mathbf{x}_t, \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \eta(\mathbf{x}_t, \boldsymbol{\theta})}{\partial \theta_j} \right]$$

risultando

$$\begin{aligned} Var(\hat{\theta}) &= \hat{\sigma}^2 \left[\sum_{t=1}^2 \left(\frac{\partial \eta(X_t, \theta)}{\partial \theta} \right)^2 \right]_{\theta=\hat{\theta}}^{-1} = 2.93336 \left\{ [(\log 2) 2^{2.05369}]^2 + \right. \\ &\quad \left. + [(\log 3) 3^{2.05369}]^2 \right\}^{-1} = \frac{2.93336}{118.2849} = 0.024799 \end{aligned}$$

Abbiamo che $\mathbf{F}_t = (\log X_t) X_t^\theta$, di conseguenza quindi $\mathbf{F}_1 = (\log 2) 2^{2.0537} = 2.8777$ e $\mathbf{F}_2 = (\log 3) 3^{2.0537} = 10.4883$. Allora $\sum \mathbf{F}\mathbf{F}' = 118.286$ e $(\sum \mathbf{F}\mathbf{F}')^{-1} = 0.008454$.

Dal momento che $\mathbf{H}_t = (\log X_t)^2 X_t^\theta$, $\mathbf{H}_1 = 1.9947$ e $\mathbf{H}_2 = 11.523$, quindi

$$\begin{aligned} \sum \mathbf{F}_t \text{tr} \left[(\sum \mathbf{F}\mathbf{F}')^{-1} \mathbf{H}_t \right] &= [2.8777(1.9947) + 10.4883(11.523)](0.008454) = \\ &= 1.0702 \end{aligned}$$

da cui

$$\text{Bias}(\hat{\theta}) = \frac{-2.9334(0.008454)(1.0702)}{2} = -0.01327$$

che in termini di percentage bias equivale a

$$\% \text{Bias}(\hat{\theta}) = \frac{-0.01327}{2.0537} 100 = -0.646\%$$

Abbiamo visto che il modello in questione ha un comportamento non lineare, mentre la regola operativa dell'1% indicherebbe qui una prossimità alla linearità del parametro ($\% \text{Bias}(\hat{\theta}) < 1\%$), questo è probabilmente ascrivibile alla poca precisione della stima causata dalla ridotta dimensione campionaria, come vedremo in seguito tale regola si rivela valida in ogni altra situazione con un n più elevato.

Valutiamo ora l'impatto della riparametrizzazione $\phi = g(\theta) = \exp(\theta)$, obiettivo per il quale possiamo servirci della formula

$$\text{Bias}(\hat{\phi}) = \mathbf{g}' \text{Bias}(\hat{\theta}) + \frac{1}{2} \text{tr} \left[\mathbf{M}\mathbf{Cov}(\hat{\theta}) \right]$$

dove i simboli hanno i significati noti, che nel nostro caso si riduce semplicemente a

$$Bias(\hat{\phi}) = Bias(\hat{\theta}) \frac{\partial g(\theta)}{\partial \theta} + \frac{1}{2} Var(\hat{\theta}) \frac{\partial^2 g(\theta)}{\partial \theta^2}$$

E' anche possibile valutare la matrice di varianze e covarianze (in questo caso semplicemente la varianza) di $\hat{\phi}$ tramite la relazione

$$Var(\hat{\phi}) = tr \left[(\mathbf{g}'\mathbf{g}) Cov(\hat{\boldsymbol{\theta}}) \right]$$

che nel nostro caso si riduce a

$$Var(\hat{\phi}) = Var(\hat{\theta}) \left[\frac{\partial g(\theta)}{\partial \theta} \right]_{\theta=\hat{\theta}}^2$$

Applicando le suddette formule otteniamo quindi che

$$\begin{aligned} \hat{\phi} &= g(\hat{\theta}) = e^{2.0537} = 7.7966 \\ \frac{\partial g(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}} &= \frac{\partial^2 g(\theta)}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} = e^{2.0537} = 7.7966 \end{aligned}$$

Dunque

$$\begin{aligned} Bias(\hat{\phi}) &= -0.01327(7.7966) + \frac{1}{2}(0.02480)7.7966 = \\ &= -0.006789 \end{aligned}$$

equivalente ad un percentage bias di

$$\% Bias(\hat{\phi}) = \frac{-0.006789}{7.7966} 100 = -0.087\%$$

Dunque la distorsione di $\hat{\phi}$ è più ridotta non solo in senso assoluto, ma anche in senso relativo, in linea con le considerazioni effettuate a riguardo dell'impatto della misura di curvatura **PE**.

La stima della varianza asintotica di $\hat{\phi}$ è, infine, pari a

$$Var(\hat{\phi}) = (0.02480)(7.7966)^2 = 1.5075$$

Tale misura, congiuntamente con la varianza asintotica di $\hat{\theta}$, risulterà utile nella seguente simulazione per effettuare una valutazione di quanto tali stime risultino avere una varianza più o meno prossima al limite inferiore costituito dalla varianza asintotica.

In quanto segue vengono riportati i risultati di studi di simulazione relativi alla distribuzione dei parametri θ e ϕ ; il listato di codice in ambiente *R* scritto per lo scopo è riportato in appendice, e parimenti lo saranno quelli per i modelli che analizzeremo in seguito. Un totale di 2000 stime per il parametro θ sono state ricavate a fronte di un insieme di 2000 risposte simulate sulla base del modello con $\hat{\theta} = 2.0537$ ed errori supposti normali di media nulla e varianza pari a $\hat{\sigma}^2 = 2.9334$; sulla base di ognuna di queste risposte si è applicato il criterio di stima dei LS per l'ottenimento del valore di $\hat{\theta}$, e si propone la rappresentazione grafica della distribuzione risultante, con sovrapposta una curva di densità normale, ed il qq-plot con riferimento ad una normale. Misure di asimmetria ed excess kurtosis sono state implementate nel codice di modo da poter avere un metro di valutazione della prossimità o meno della distribuzione a quella normale, inoltre si è implementato il calcolo di due misure di percentage bias e di excess variance come segue:

$$\%Bias(\hat{\theta}) = \frac{E(\hat{\theta}) - \theta}{\theta} 100$$

$$\%Var(\hat{\theta}) = \frac{Var(\hat{\theta}) - Var(\theta)}{Var(\theta)} 100$$

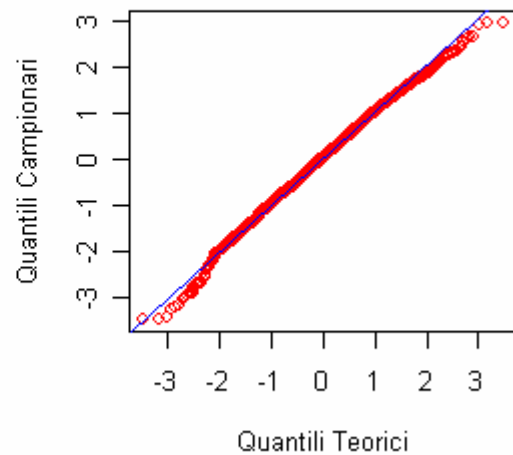
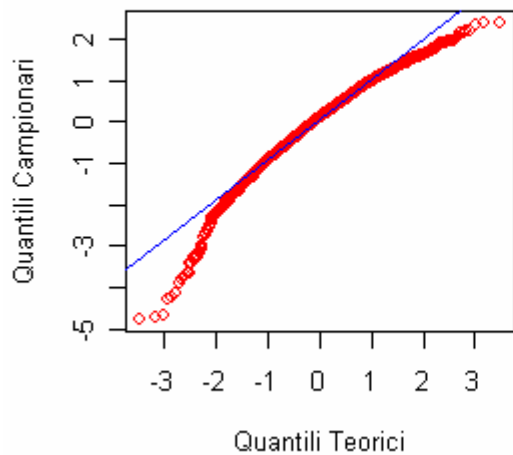
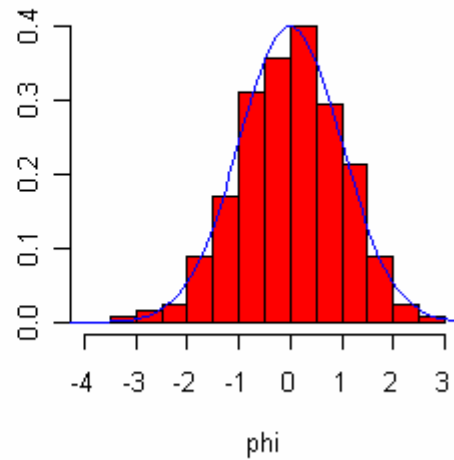
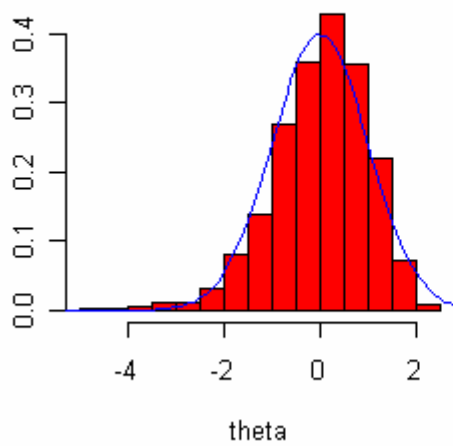
e dove, ai fini della stima, a θ e $Var(\theta)$ sono stati sostituiti, rispettivamente, i valori $\hat{\theta} = 2.0537$ e $\hat{\sigma}^2 = 2.9334$. La misura di excess variance fornisce una indicazione di quanto, in percentuale, la stima parametrica si discosti dal limite inferiore di varianza, individuato dalla matrice asintotica di varianze e covarianze (in questo caso solo la varianza).

Identica procedura è stata portata in essere nei confronti del parametro ϕ , per il quale non è stata necessaria una nuova simulazione, bensì si è proceduto alla trasformazione dell'insieme di stime ottenute per θ tramite la relazione $\phi = \exp(\theta)$ e si sono valutate le medesime misure impiegate per il parametro originario.

Le rappresentazioni grafiche delle distribuzioni delle stime parametriche sono relative alla normalizzazione delle stesse sottraendo la media e dividendo per lo scarto quadratico medio, di modo da ottenere distribuzioni di media nulla e varianza unitaria.

L'esecuzione della funzione implementata per le simulazioni porta ai seguenti risultati ed ai grafici presentati nella prossima pagina:

```
[1] "Parametro theta"
[1] "Skewness           :" "-0.787612049237392"
[1] "Excess Kurtosis    :" "1.40417905692632"
[1] "% Bias             :" "-0.663568769490854"
[1] "% Excess Variance  :" "11.8116539180589"
[1] "Parametro phi"
[1] "Skewness           :" "-0.19227954018488"
[1] "Excess Kurtosis    :" "0.080993447115771"
[1] "% Bias             :" "-0.0333560709235949"
[1] "% Excess Variance  :" "0.47318743733441"
```

La distribuzione relativa al parametro θ risulta avere una asimmetria negativa (pari a -0.7876), visibile anche graficamente, più marcata rispetto a quella del parametro ϕ (pari a -0.1923), così come una leptocurtosi molto più accentuata (1.4042 contro 0.08099 di ϕ). La formula per la stima della distorsione di Box ha previsto sufficientemente bene il bias nelle due distribuzioni, avendo fornito un valore per θ pari a -0.646% , a fronte di un effettivo -0.664% , e per ϕ pari a -0.087% contro un risultato di simulazione di -0.0334% . L'indicatore di excess variance risulta essere pari all' 11.81% per θ e allo 0.47% per ϕ , mettendo in luce quindi una minore variabilità della stima di ϕ in termini di lontananza dal limite inferiore. Potremmo

quindi considerarla più efficiente in questo senso. L'analisi dei grafici quantile-quantile di probabilità normale evidenzia l'asimmetria più marcata della distribuzione di θ (forma a "U") e corrobora l'evidenza di una distribuzione più prossima ad una gaussiana per ϕ . In questi termini abbiamo verificato empiricamente l'impatto, prima affermato solo in via teorica, della non linearità **PE** sulle proprietà degli stimatori di un semplice modello non lineare, e mostrato come sia possibile ridurla tramite un'opportuna riparametrizzazione.

Una considerazione va portata riguardo a quale possa essere la riparametrizzazione corretta. Le misure e le considerazioni introdotte aiutano a scegliere verso quale o quali parametri orientare la riparametrizzazione, ma non forniscono alcuna indicazione su quale possa essere quella opportuna; questo compito ricade sulle spalle della teoria sottostante il problema in esame, in particolar modo sulla interpretabilità o meno del parametro una volta trasformato. In molte applicazioni, ad esempio in ambito chimico, i parametri hanno un ben determinato significato fisico, e trasformazioni arbitrarie potrebbero stravolgerlo; in alcuni casi può non essere un problema così grande, si considerino grandezze come il decibel o il pH, che sono per definizione trasformate logaritmiche di altre quantità, ed una eventuale riparametrizzazione a mezzo di un esponenziale porterebbe comunque a parametri interpretabili. In linea generale però questo non è garantito, ed è necessario usare quindi molta cautela nella riparametrizzazione; supposto che tale procedura non implichi eccessivi problemi nei termini appena definiti, una regola operativa potrebbe riguardare la forma della distribuzione delle stime a seguito di una simulazione: qualora la distribuzione si presentasse asimmetrica negativamente una trasformata esponenziale potrebbe ricondurla ad una forma più simmetrica, e per converso a fronte di una asimmetria positiva una trasformata logaritmica potrebbe portare ad ottenere il medesimo risultato.

Il modello a due parametri di Michaelis-Menten

La combinazione che segue è tratta da Bates e Watts (1988) e si riferisce al modello di Michaelis-Menten, molto impiegato in ambito chimico nello studio della velocità delle reazioni enzimatiche. L'esperimento in questione riguarda la relazione fra la velocità (y) di una reazione enzimatica e la concentrazione di substrato (x), esperimento in cui l'enzima è stato trattato con Puromycina. I parametri hanno la seguente interpretazione fisica: θ_1 indica la concentrazione massimale asintotica quando $\theta_2 > 0$, e θ_2 è la concentrazione a metà della velocità massimale; ambedue i parametri quindi dovranno essere positivi perché il fenomeno sia reale.

$$\text{Modello: } Y_t = \frac{\theta_1 X_t}{\theta_2 + X_t} + \varepsilon_t \Rightarrow \eta(\boldsymbol{\theta}) = E(Y) = \frac{\theta_1 X}{\theta_2 + X}$$

Dati:

X_t	Y_t
0.02	76
0.02	47
0.06	97
0.06	107
0.11	123
0.11	139
0.22	159
0.22	152
0.56	191
0.56	201
1.10	207
1.10	200

Come si può vedere l'esperimento è stato completamente replicato una volta, questo ci consentirà di trattare le seguenti diverse situazioni:

1. Si hanno a disposizione solo $n = 3$ osservazioni
2. Si ha a disposizione il disegno completo, $n = 6$, ma senza repliche
3. Si ha a disposizione il disegno completo replicato una volta, $n = 12$

Potremo così valutare l'impatto sulla non linearità tanto dell'aumentata dimensione campionaria quanto l'effetto della replicazione del disegno sperimentale.

Caso 1. $n = 3$ osservazioni

Supponendo di disporre solo di una parte del disegno sperimentale analizziamo il comportamento del modello e delle stime.

X_t	Y_t
0.02	76
0.22	152
1.10	200

R fornisce come stima secondo i LS il vettore $\hat{\theta} = (194.99256277, 0.03623891)'$ e l'esecuzione della funzione *rms.curv* fornisce i seguenti valori di non linearità intrinseca ed indotta dai parametri:

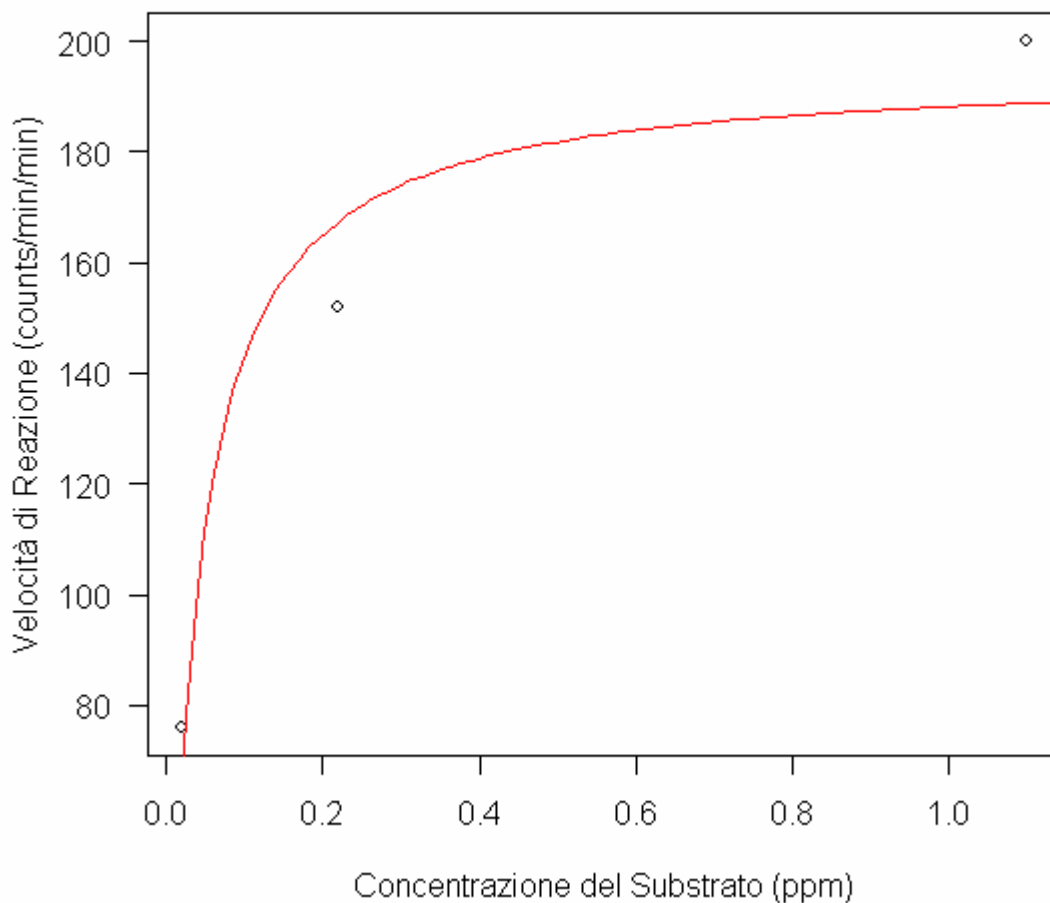
Parameter effects: $c^{\theta} \times \sqrt{F} = 6.9973$

Intrinsic: $c^{\iota} \times \sqrt{F} = 2.1581$

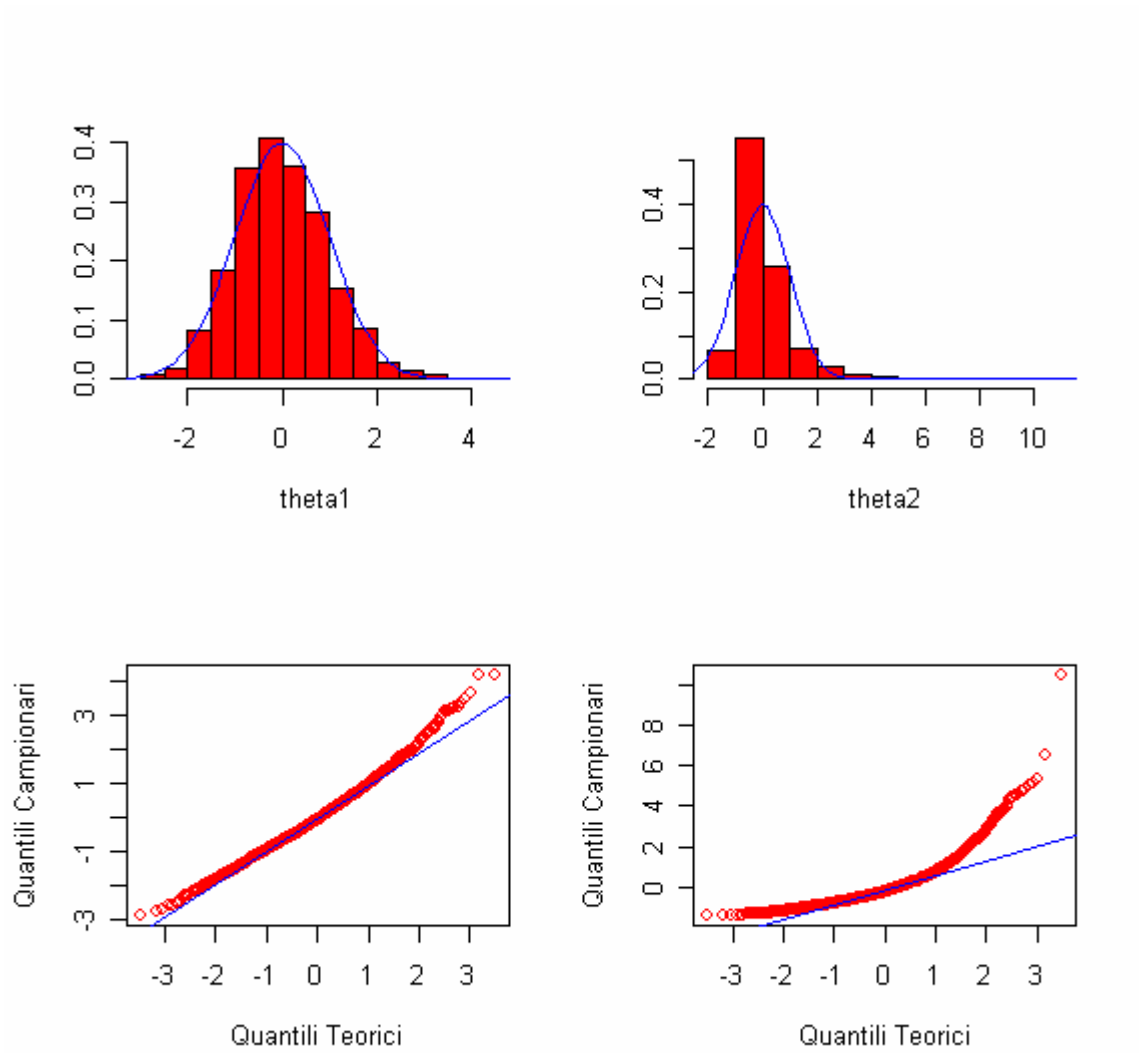
rms.curv fornisce come output i valori di trasformate delle curvature (d'ora in poi faremo riferimento a questi), denominate *Root Mean Square Curvatures*, le quali rappresentano la media potenziata di ordine 2 delle curvature lungo tutte le

direzioni, moltiplicate per \sqrt{F} , con $F = F(2,1;0.05) = 199.5$. Furono gli stessi Bates e Watts ad introdurre questa trasformazione delle iniziali misure definite in questa tesi, ed indicano come soglia critica un valore non in eccesso di 0.3 per l'accettabilità delle assunzioni planare e di coordinate uniformi; di conseguenza la conclusione che possiamo trarre è che ambedue le curvature risultano essere molto elevate.

Modello di Michaelis Menten adattato ai dati con n=3



Uno studio tramite simulazioni, analogamente a quanto fatto con riguardo al modello ad un parametro, è stato effettuato; i risultati delle statistiche relative a simmetria, excess kurtosis, percentage bias ed excess variance sono riportati nella seguente pagina, insieme ai grafici di dette simulazioni ed al grafico di probabilità normale relativo ad ogni stima.



[1] "Parametro theta1"

[1] "Skewness : " "0.414831766142678"

[1] "Excess Kurtosis : " "0.479846092469153"

[1] "% Bias : " "1.27639969714024"

[1] "% Excess Variance : " "46031.9361612799"

[1] "Parametro theta2"

[1] "Skewness : " "2.33185074149329"

[1] "Excess Kurtosis : " "10.5750414330195"

[1] "% Bias : " "16.3517347305427"

[1] "% Excess Variance : " "78773.0651718388"

La stima della distorsione percentuale secondo Box, il codice per la quale è presente in appendice, fornisce come risultati il valore di $-5.43e-07$ per θ_1 e 8.292 per θ_2 ; tale

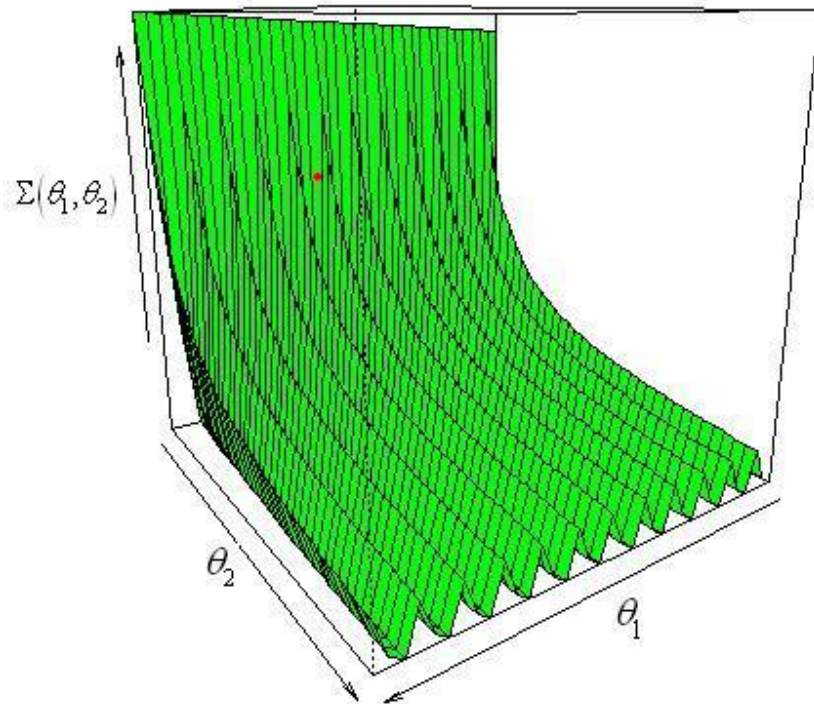
stima non è molto precisa confrontata con questa simulazione, probabilmente a causa della ridotta dimensione campionaria, riuscendo però comunque a rilevare la maggior distorsione del secondo parametro rispetto al primo.

È visibile (e rilevabile dai valori di skewness) una asimmetria positiva delle distribuzioni, particolarmente marcata per il secondo parametro; θ_1 possiede una distribuzione più prossima alla normale, mentre quella di θ_2 ne risulta evidentemente molto lontana. Risulta particolarmente evidente inoltre la bassa efficienza dal dato della excess variance, si sono ottenuti i due spropositati valori di 4631.94% nel primo caso e 78773.06% nel secondo. È auspicabile che riducendo la non linearità indotta dai parametri questa situazione tenda a migliorare, ma non avendo particolari indicazioni derivanti dai grafici o dalla teoria sottostante riguardo ad eventuali convenienti riparametrizzazioni la via da seguire non è chiara. Vedremo nei casi 2. e 3. l'analisi relativamente ad una dimensione campionaria maggiore e la sua replicazione.

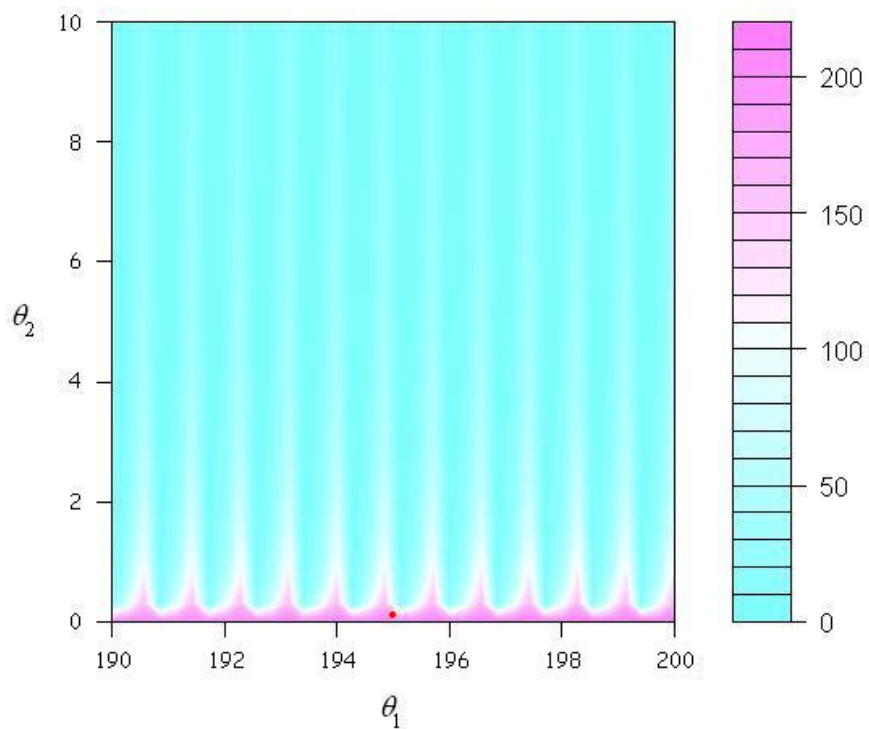
In ultimo, trovandoci nel caso $n=3$ e $p=2$ possiamo visualizzare graficamente il luogo delle soluzioni del modello come una varietà riemanniana bidimensionale (una superficie curva) immersa in uno spazio ambiente euclideo tridimensionale. Dalle considerazioni riguardo alla natura dei parametri possiamo trarre la conclusione che la zona di interesse su questa superficie sarà quella corrispondente ai valori positivi di ambedue i parametri. Per lo scopo è stato impiegato il comando *persp* in ambiente *R*, tale funzione riproduce graficamente una matrice i cui elementi corrispondono ai valori assunti lungo il terzo asse (ovvero il valore del luogo delle soluzioni) calcolati a fronte di un determinato intervallo dei restanti due assi (ovvero quelli individuati dai due parametri). Il grafico risultante per la zona di interesse è presentato nella pagina seguente congiuntamente ad un grafico relativo alle curve di livello della funzione del luogo delle soluzioni, per avere meglio visuale della natura di superficie curva dello stesso. Si è indicato in rosso il punto sul luogo delle soluzioni individuato dalla stima parametrica, e si può osservare come un intorno della sua posizione su tale superficie individui una zona fortemente ondulata; di conseguenza l'area immediatamente circostante non è sufficientemente bene approssimabile dal piano tangente nel punto (come indica la misura di curvatura

IN), mentre le curve coordinate (non rappresentate) presenteranno un comportamento poco regolare (come suggerito dalla misura di curvatura PE).

Luogo delle Soluzioni per $190 < \theta_1 < 200$
 $0 < \theta_2 < 10$



Curve di livello del Luogo delle Soluzioni

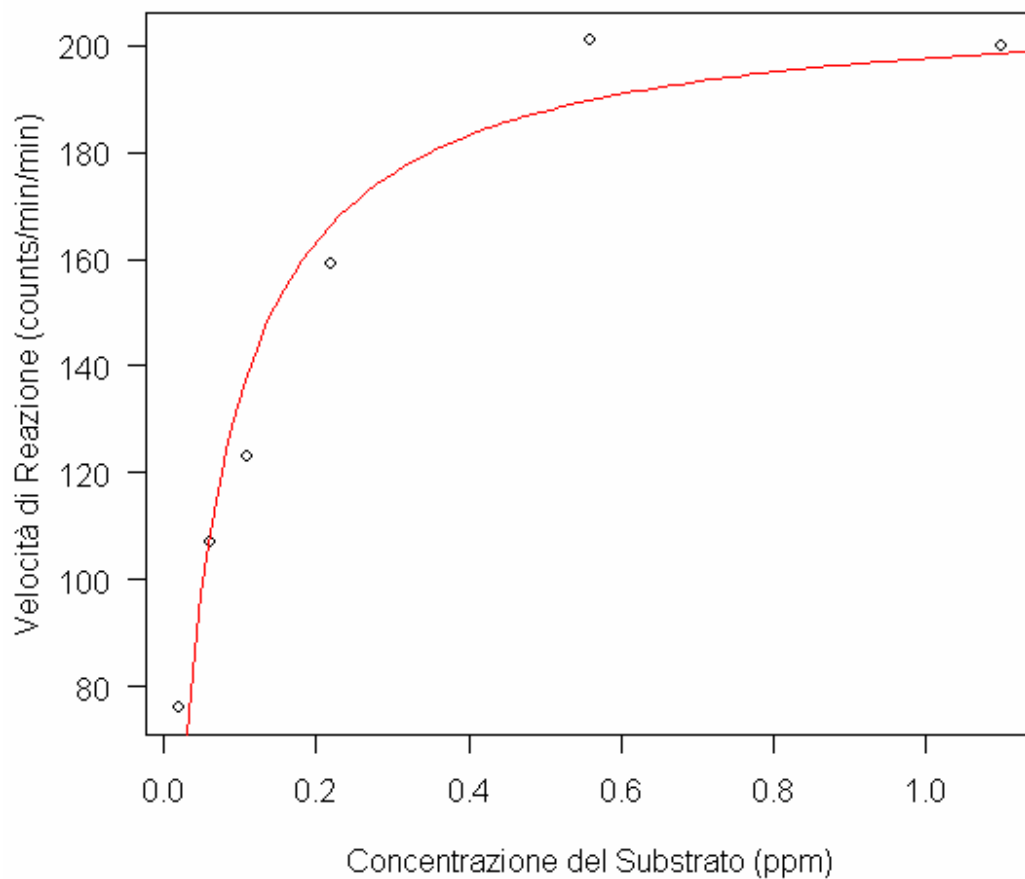


Caso 2. $n = 6$, disegno completo senza replicazioni

X_t	Y_t
0.02	76
0.06	107
0.11	123
0.22	159
0.56	201
1.10	200

La scelta di quali risposte selezionare dal data-set originario è stata arbitraria. Le stime parametriche per questo secondo disegno risultano essere $\hat{\theta} = (208.57039415, 0.05550961)'$.

Modello di Michaelis Menten adattato ai dati con $n=6$



Le misure di curvatura stimate da *rms.curv* sono

Parameter effects : $c^{\theta} \times \sqrt{F} = 0.5212$

Intrinsic : $c^{\iota} \times \sqrt{F} = 0.2334$

Risulta immediatamente evidente la drastica riduzione degli effetti di curvatura intrinseca ed indotta dai parametri. L'aggiunta di punti sul piano sperimentale ha permesso di ridurre la non linearità del luogo delle soluzioni; geometricamente possiamo ritenere questo fatto dovuto al maggior grado di precisione con cui si è definita la superficie, permettendo così una approssimazione migliore tramite il piano tangente ed una visuale più "vicina" tale da rendere meno percepibile la non uniformità delle linee coordinate. Nonostante questo la curvatura **PE** risulta essere ancora rilevante rispetto alla soglia indicativa di 0.3.

Per verificare l'impatto di questa diminuzione di curvatura sulle proprietà delle stime si è proceduto ad una ulteriore simulazione sulla base del nuovo disegno. I risultati sono i seguenti:

[1] "Parametro theta1"

[1] "Skewness : " "0.174794240074771"

[1] "Excess Kurtosis : " "-0.122365978932609"

[1] "% Bias : " "0.211080120444712"

[1] "% Excess Variance : " "21867.2104661784"

[1] "Parametro theta2"

[1] "Skewness : " "0.568915606829640"

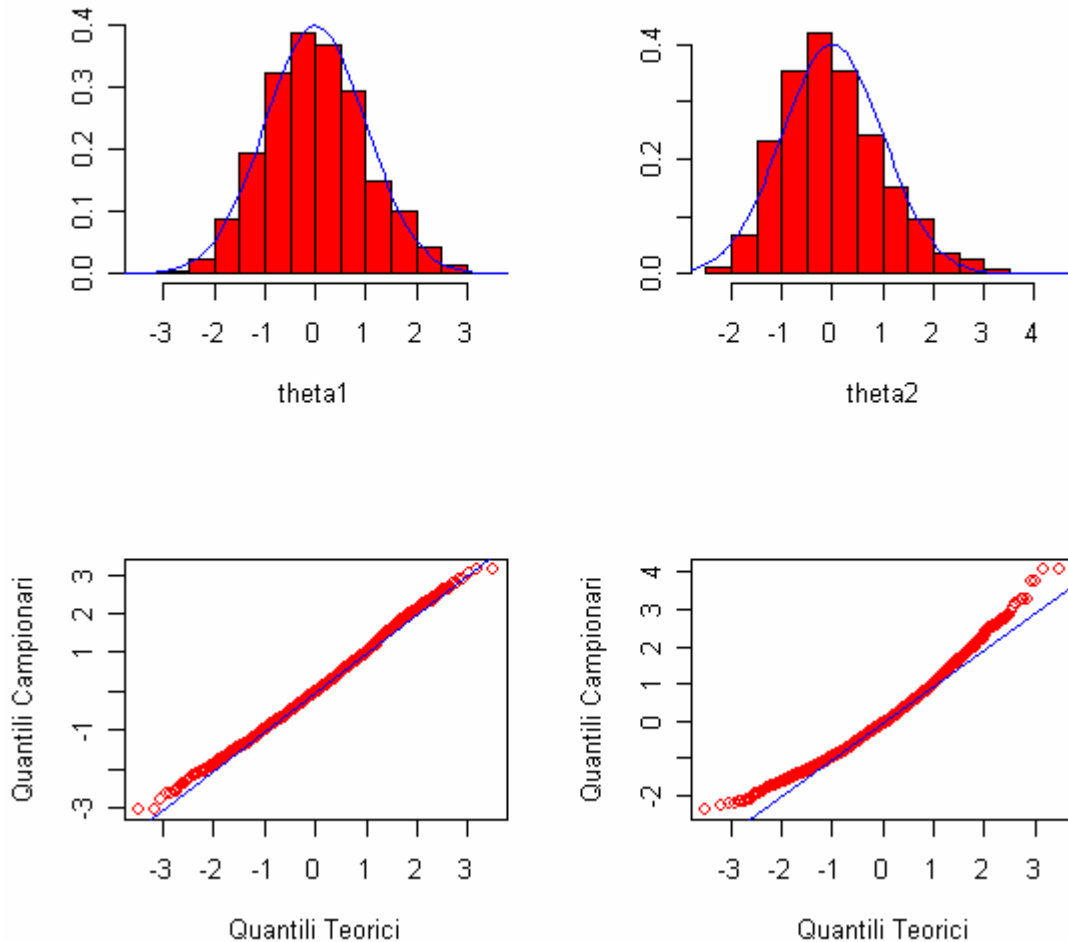
[1] "Excess Kurtosis : " "0.341765846220973"

[1] "% Bias : " "2.71289263227098"

[1] "% Excess Variance : " "22362.8928631632"

La stima del percentage bias tramite la formula di Box fornisce i valori di $-1.9e-08$ per θ_1 e 1.168549 per θ_2 ; notiamo anche in questo caso come la valutazione sia non

accuratissima rispetto alla simulazione, ma comunque consente l'individuazione della stima più distante dalla linearità grazie alla regola operativa dell'1%.



L'analisi dei grafici derivanti dalla simulazione rende visivamente più evidente il beneficio dell'aumento della dimensione campionaria: la stima di θ_2 , in precedenza ben lungi dal poter essere considerata approssimabile ad una normale, presenta ora una asimmetria positiva più contenuta e solo una lieve leptocurtosi; la distribuzione di θ_1 , già non troppo lontana dalla gaussianità, ne risulta ulteriormente migliorata in questo senso: l'asimmetria e l'excess kurtosis, così come il percentage bias, risultano sensibilmente diminuiti. Per quanto riguarda l'excess variance possiamo notarne l'enorme decremento rispetto al Caso 1., pur restando comunque su livelli molto elevati, indice della poca efficienza delle stime.

Caso 3. $n = 12$, replicazione del disegno di cui al Caso 2.

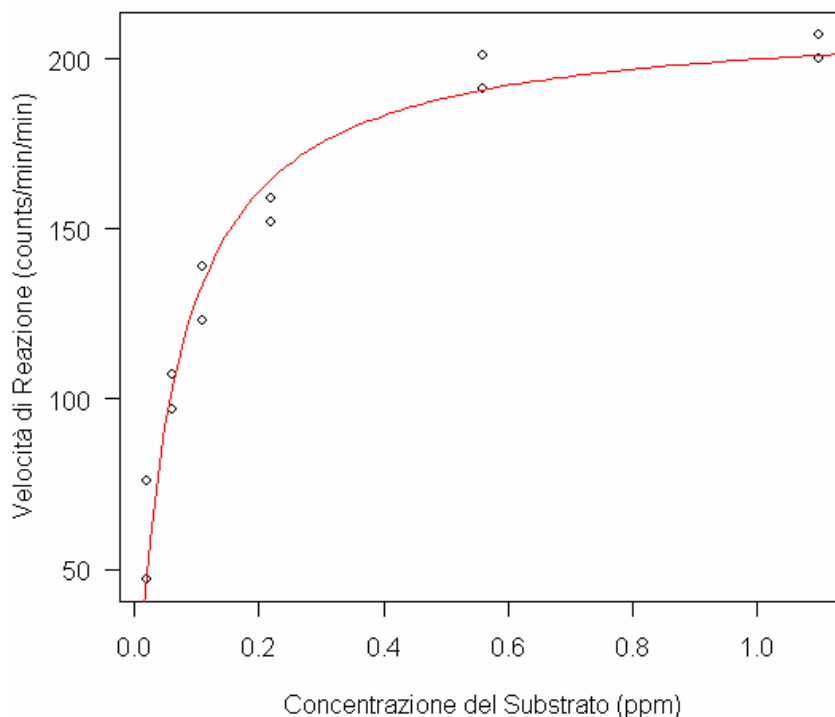
Consideriamo ora il disegno sperimentale nel suo complesso così come presentato nella tabella iniziale. La stima parametrica porta a $\hat{\theta} = (212.6836298, 0.0641211)'$, i valori di percentage bias stimati dalla formula di Box sono $-6.29e-08$ per θ_1 e 0.308 per θ_2 . Le misure di curvatura risultano essere

Parameter effects : $c^{\theta} \times \sqrt{F} = 0.2121$

Intrinsic : $c^{\iota} \times \sqrt{F} = 0.092$

da cui possiamo rilevare l'ulteriore decremento della non linearità nell'intorno del punto di stima parametrica sul luogo delle soluzioni, decremento atteso in virtù delle considerazioni teoriche sull'impatto di una replicazione completa del disegno sperimentale portate nel paragrafo 4.3. Ci attenderemo un miglioramento nelle proprietà delle stime, miglioramento che provvederemo a verificare tramite una simulazione come nei casi precedenti.

**Modello di Michaelis Menten adattato ai dati con $n=12$
Replicazione del Caso 2.**



La simulazione ha portato ai seguenti risultati:

[1] "Parametro theta1"

[1] "Skewness : " "0.0586674668890446"

[1] "Excess Kurtosis : " "-0.0356990046071877"

[1] "% Bias : " "0.132904314742775"

[1] "% Excess Variance : " "12310.2901730858"

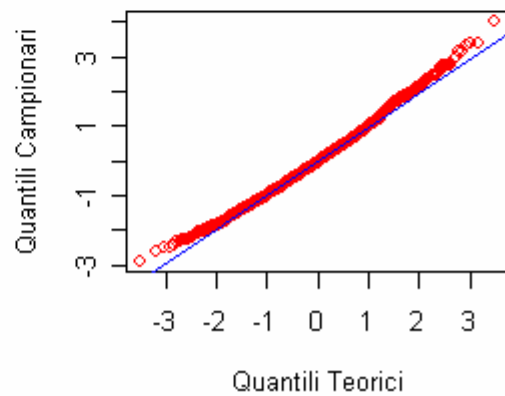
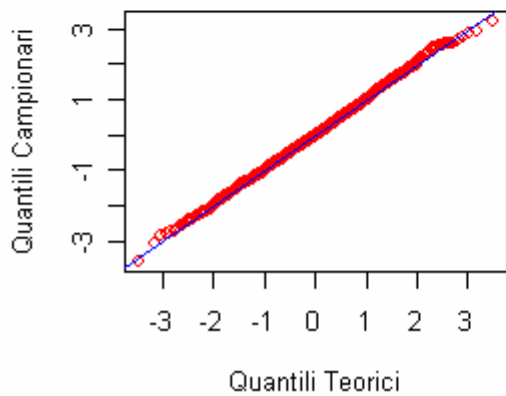
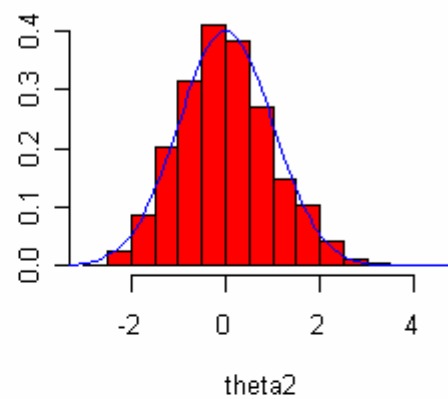
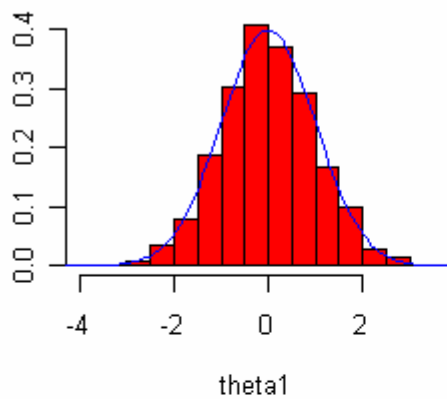
[1] "Parametro theta2"

[1] "Skewness : " "0.310108173817681"

[1] "Excess Kurtosis : " "0.0759483952330715"

[1] "% Bias : " "0.824803472543305"

[1] "% Excess Variance : " "12655.1437968884"



La replicazione del disegno sperimentale ha portato ad un miglioramento in termini di asimmetria, excess kurtosis e distorsione per entrambe le distribuzioni, ferma restando una più accentuata (per quanto lieve) asimmetria positiva per θ_2 . Anche l'excess variance risulta sensibilmente diminuita, per quanto ancora molto elevata per le due stime; in definitiva si può affermare che, per la combinazione modello-data set considerata, l'aumento della dimensione campionaria e la replicazione del piano sperimentale hanno apportato visibili benefici alle proprietà delle stime, come mostrato.

Un modello di tipo Weibull a quattro parametri

La combinazione modello-data set seguente è tratta da Ratkowsky (1983). I dati si riferiscono alla resa Y di una coltura di una particolare specie di cipolle, e si desidera interpretare tale resa in relazione al tempo X , a fronte di un modello a quattro parametri di tipo Weibull ampiamente impiegato per lo studio di questo genere di fenomeni.

$$\text{Modello: } Y_i = \theta_1 - \theta_2 \exp(-\theta_3 X_i^{\theta_4}) + \varepsilon_i \Rightarrow \eta(\boldsymbol{\theta}) = E(Y) = \theta_1 - \theta_2 \exp(-\theta_3 X^{\theta_4})$$

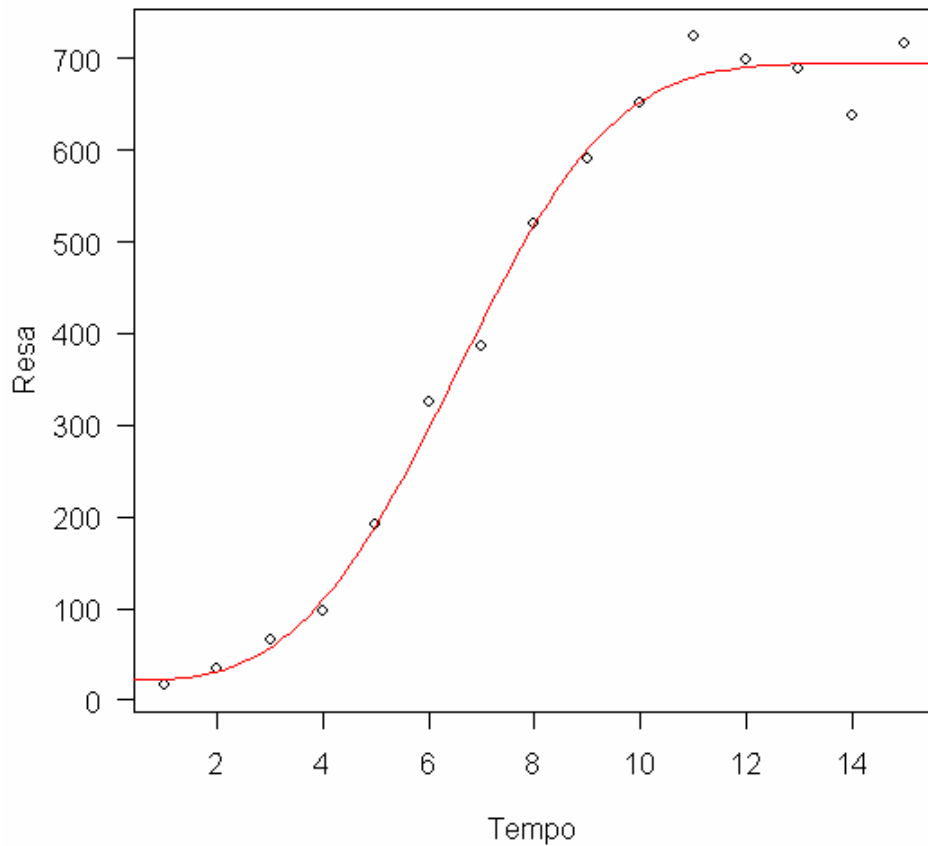
Dati:

X_i	Y_i
1	16.08
2	33.83
3	65.80
4	97.20
5	191.55
6	326.20
7	386.87
8	520.53
9	590.03
10	651.92
11	724.93
12	699.56
13	689.96
14	637.56
15	717.41

La procedura di stima fornisce il vettore parametrico $\hat{\boldsymbol{\theta}} = (695, 673.5, 0.00152, 3.262)'$

ed il grafico del modello adattato è il seguente:

Modello di tipo Weibull



Le misure di curvatura risultano essere

Parameter effects : $c^{\theta} \times \sqrt{F} = 10.6246$

Intrinsic : $c^{\iota} \times \sqrt{F} = 0.1969$

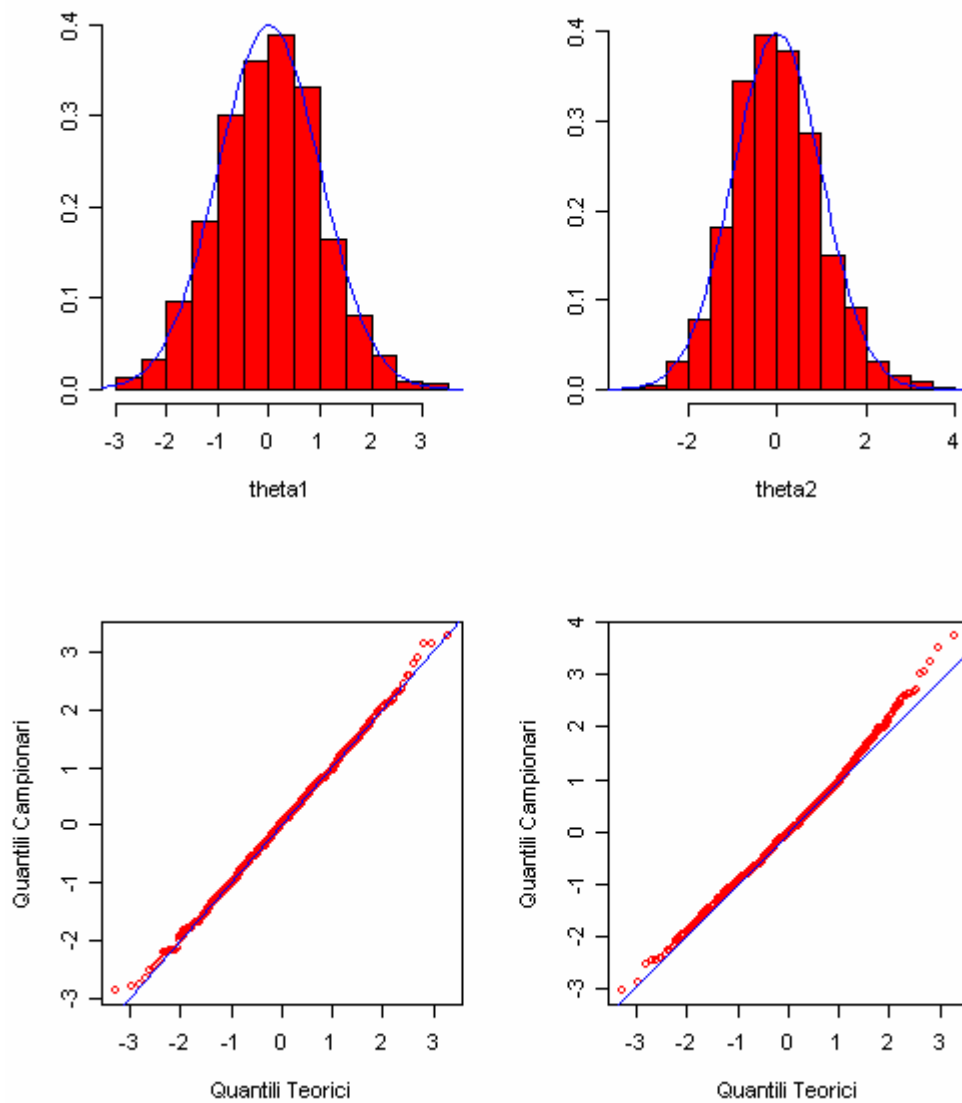
indicando una accettabile non linearità intrinseca ed una elevatissima non linearità indotta dalla scelta parametrica. È stata effettuata una simulazione per verificare empiricamente le proprietà delle stime tramite gli indicatori già impiegati negli esempi precedenti, i risultati sono riportati nella pagina seguente. La misura di percentage bias è risultata essere inferiore all'1% per tutti i parametri eccetto che per θ_3 , registrando un valore dell'oltre il 16%, il che suggerisce che l'elevata curvatura PE sia causata principalmente (se non quasi unicamente) da tale parametro.


```

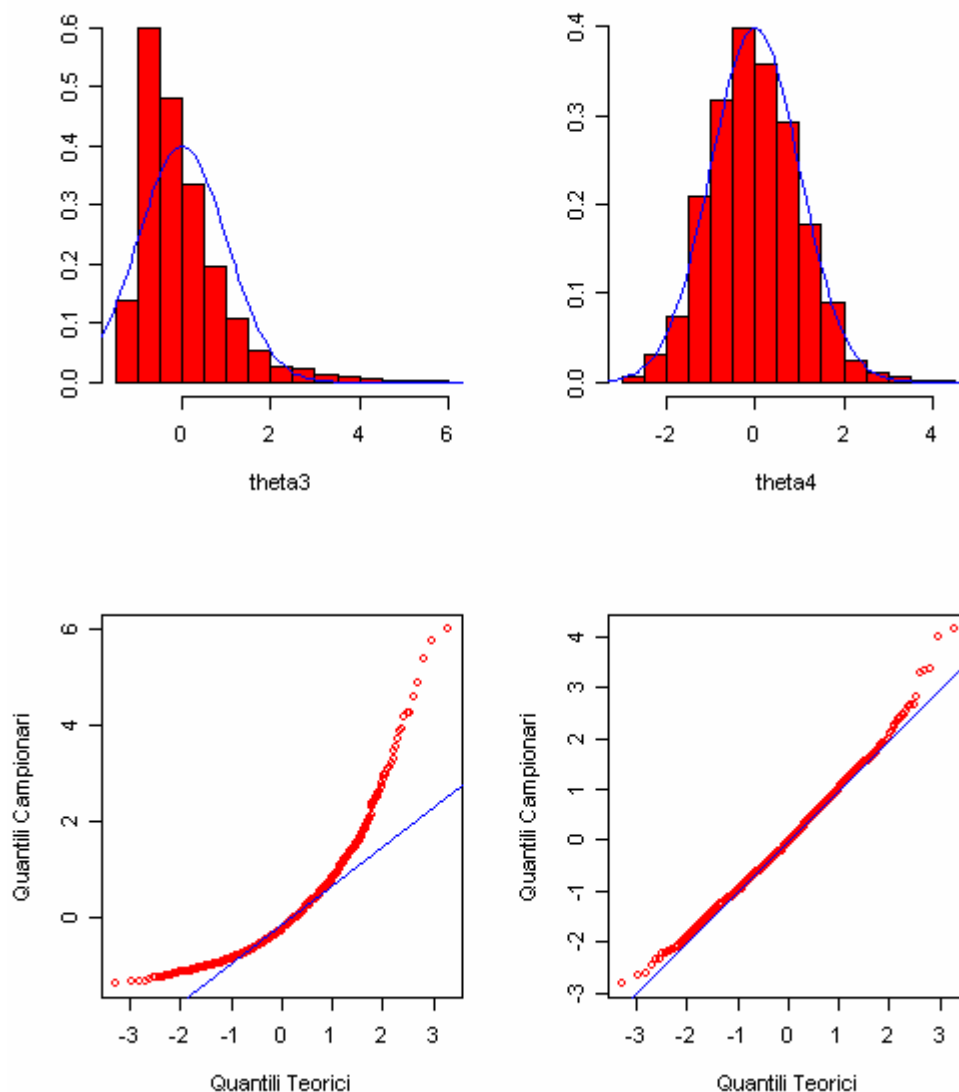
[1] "Parametro theta1"
[1] "Skewness          		:" "0.0424174679957538"
[1] "Excess Kurtosis    	:" "-0.0230925580759545"
[1] "% Bias             		:" "0.0315410909000957"
[1] "% Excess Variance  	:" "72463.2646870802"
[1] "Parametro theta2"
[1] "Skewness          		:" "0.266069498445078"
[1] "Excess Kurtosis    	:" "0.295886660063375"
[1] "% Bias             		:" "0.120332251436929"
[1] "% Excess Variance  	:" "67683.2303115508"
[1] "Parametro theta3"
[1] "Skewness          		:" "1.89286791030867"
[1] "Excess Kurtosis    	:" "5.43765805702144"
[1] "% Bias             		:" "20.4540279609399"
[1] "% Excess Variance  	:" "114989.080299571"
[1] "Parametro theta4"
[1] "Skewness          		:" "0.286070669394756"
[1] "Excess Kurtosis    	:" "0.330023790514068"
[1] "% Bias             		:" "0.623673845055484"
[1] "% Excess Variance  	:" "76174.5435009961"

```

Risulta subito evidente come il comportamento più spiccatamente non lineare sia proprio quello di θ_3 : marcata asimmetria ed ancora più accentuata leptocurtosi, una distorsione percentuale del 20% (in linea con la stima secondo la formula di Box) ed una excess variance che supera quelle di ogni altro parametro (per quanto comunque molto elevate). I grafici che seguono sono relativi a questa simulazione, e forniscono una ulteriore riprova visiva delle considerazioni portate.



Le distribuzioni di θ_1 e θ_2 presentano una asimmetria ed una curtosi molto contenute (in special modo θ_1) e non vi è evidenza di un comportamento che si discosti in modo spiccato dalla gaussianità, come mostrano anche i grafici di probabilità normale.



Per quanto riguarda la distribuzione di θ_4 possono replicarsi le considerazioni portate nei confronti delle prime due, la normalità distributiva è molto avvicinata, seppur leggermente meno che nei casi di θ_1 e θ_2 . La situazione di θ_3 è invece visibilmente assai lontana dalla comoda ipotesi di gaussianità: la forma marcatamente ad “U” rivolta verso l’alto del grafico di probabilità normale è indice di una forte asimmetria positiva, come infatti è stato possibile verificare tramite il valore di skewness di 1.893, l’elevata leptocurtosi, subito visibile, riflette l’indicazione di un indice di excess kurtosis pari a 5.438.

Risulta dunque evidente come l'elevato valore di curvatura **PE** sia da imputarsi principalmente al comportamento fortemente non lineare di θ_3 . Fortunatamente in questo caso la rappresentazione grafica ci viene in aiuto, fornendoci una utile indicazione su una possibile riparametrizzazione da attuare: l'elevata leptocurtosi e la lunga coda destra suggeriscono una distribuzione sottostante prossima ad una lognormale, per riavvicinare alla normalità tale distribuzione prenderemo dunque in considerazione una trasformata esponenziale. Dal momento che i valori di θ_3 sono tipicamente inferiori all'unità si impiegherà la trasformata $\theta_3 = \exp(-\gamma)$ di modo da ottenere valori positivi per le stime del nuovo parametro, fatto quest'ultimo più evidente esprimendo la relazione come $\gamma = g(\theta_3) = \ln(1/\theta_3)$.

Le curvature calcolate sul modello dopo la riparametrizzazione sono le seguenti:

Parameter effects : $c^{\theta} \times \text{sqrt}(F) = 0.4652$

Intrinsic : $c^{\iota} \times \text{sqrt}(F) = 0.1969$

Si nota come la curvatura **IN** sia rimasta invariata, mentre la curvatura **PE** risulti drasticamente ridotta, essendo tuttavia ancora significativa in senso statistico.

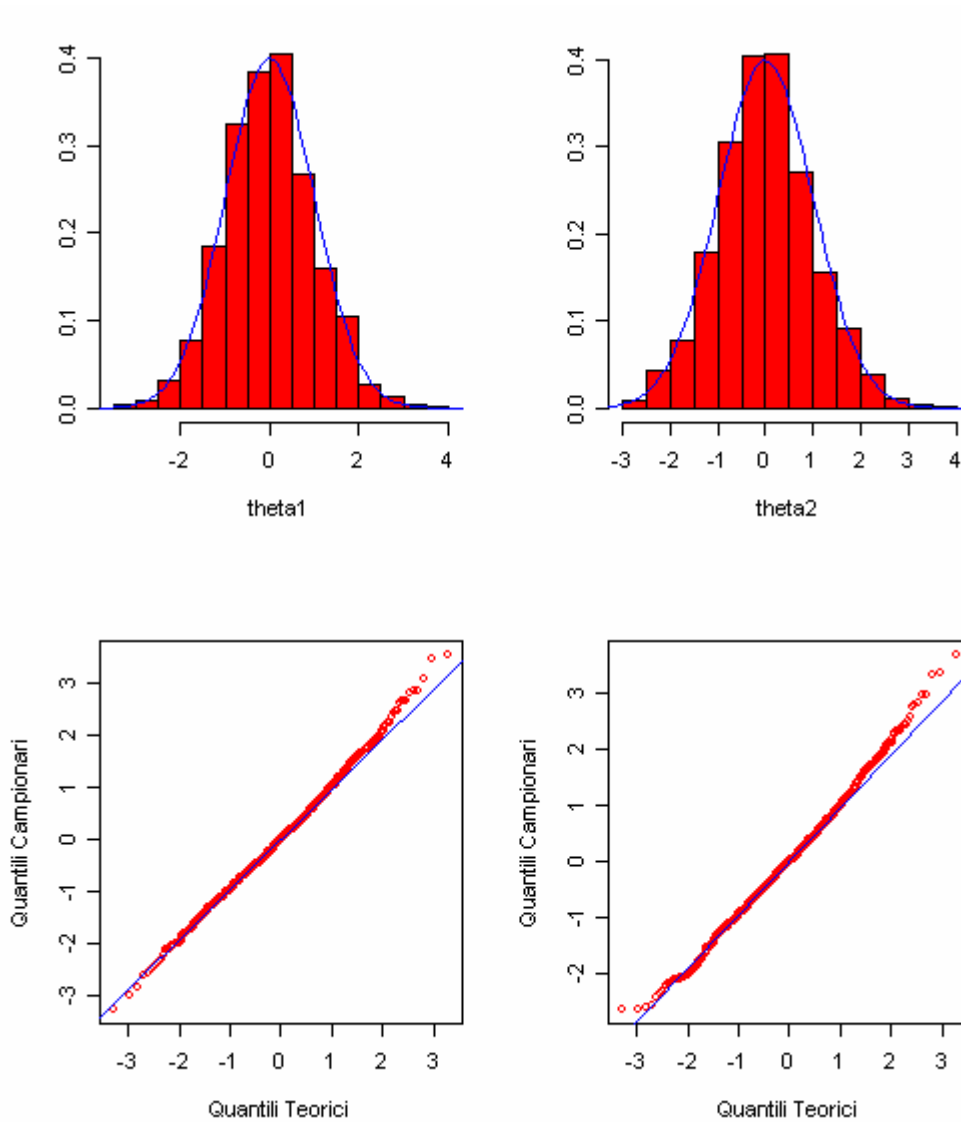
Gli effetti di tale riparametrizzazione sul percentage bias possono essere calcolati senza dover applicare nuovamente la formula di Box, sarà infatti sufficiente la relazione già sfruttata nell'esempio del modello ad un parametro, e per cui otteniamo che, a fronte di una stima $\hat{\gamma} = \ln(1/\hat{\theta}_3) = 6.48744$, la distorsione vale $\%Bias(\hat{\gamma}) = -0.917$, indubbiamente un grande guadagno rispetto alla situazione precedente. Seguono i risultati della simulazione sul modello con la nuova parametrizzazione insieme ai grafici..

```

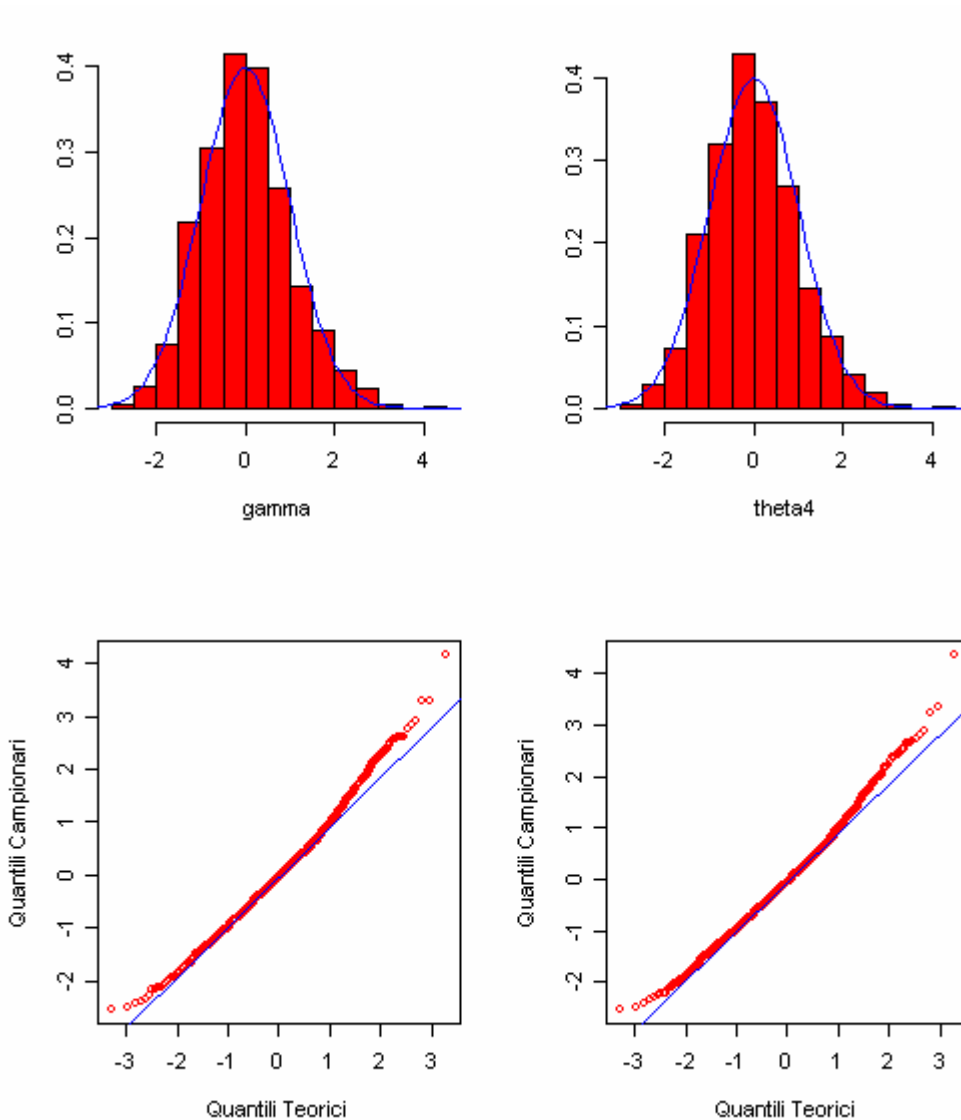
[1] "Parametro theta1"
[1] "Skewness          		:" "0.159432060516726"
[1] "Excess Kurtosis    		:" "0.213320455349691"
[1] "% Bias             		:" "0.078009852195322"
[1] "% Excess Variance  		:" "74809.5560914382"
[1] "Parametro theta2"
[1] "Skewness          		:" "0.200078702555838"
[1] "Excess Kurtosis    		:" "0.199749732557699"
[1] "% Bias             		:" "0.221992484694849"
[1] "% Excess Variance  		:" "73611.9973112389"
[1] "Parametro gamma"
[1] "Skewness          		:" "0.374060749857271"
[1] "Excess Kurtosis    		:" "0.279056131981186"
[1] "% Bias             		:" "0.872098164800406"
[1] "% Excess Variance  		:" "77779.9670241328"
[1] "Parametro theta4"
[1] "Skewness          		:" "0.383845373403495"
[1] "Excess Kurtosis    		:" "0.332925418157693"
[1] "% Bias             		:" "0.907997554196814"
[1] "% Excess Variance  		:" "78865.9553893043"

```

Dall'esame delle statistiche risulta evidente come la riparametrizzazione abbia ricondotto ampiamente il terzo parametro del modello entro limiti accettabili di prossimità alla linearità, l'asimmetria e l'excess kurtosis sono contenute, e così il bias che resta al di sotto dell'1%. Anche la excess variance, per quanto sempre elevata, risulta di molto ridotta da questa riparametrizzazione. L'impatto sul comportamento degli altri parametri di questa riparametrizzazione sembra essere non apprezzabile, come è possibile visualizzare analizzando i grafici delle simulazioni nelle pagine seguenti.



Non si evidenziano significativi cambiamenti nel comportamento distributivo di θ_1 e θ_2 , che conservano la loro già buona prossimità alla linearità, ed analoga considerazione può essere fatta anche a riguardo del parametro θ_4 , di cui il grafico nella pagina seguente.



Risulta subito evidente il netto miglioramento in termini di prossimità alla distribuzione normale per quanto riguarda il terzo parametro del modello; γ risulta distribuito in modo approssimativamente normale, con leggera asimmetria positiva e leptocurtosi. La riparametrizzazione ha dunque raggiunto lo scopo previsto di ricondurre verso la linearità il comportamento del parametro che più contribuiva alla curvatura **PE** per questa combinazione di modello-data set; tale effetto non è del tutto svanito, come testimonia il valore di **PE** ancora superiore a 0.3, e volendo sarebbe possibile continuare a riparametrizzare, coinvolgendo anche gli altri parametri, nel tentativo di migliorare ancora. Però, sebbene statisticamente questa indicazione di curvatura rimanga significativa, la discreta approssimazione alla normalità per tutti i parametri congiuntamente potrebbe indurre il ricercatore a

ritenersi soddisfatto e a non voler proseguire oltre, giudicando tale approssimazione sufficientemente buona da non invalidare l'impiego delle procedure inferenziali e delle varie tecniche statistiche che poggiano su assunzioni di normalità distributiva delle stime parametriche.

Conclusioni

Lo scopo di questa tesi è stato quello di presentare un'analisi di alcune di quelle tecniche statistiche che prendono le loro mosse dalla branca della Matematica nota come Geometria Differenziale. Dopo una breve presentazione del problema della stima non lineare secondo il principio dei minimi quadrati, e l'interpretazione geometrica di quest'ultimo con riferimento al modello di regressione lineare nel **primo capitolo**, si è resa necessaria una lunga introduzione di concetti intuitivi e definizioni formali nei **capitoli secondo e terzo** di modo da poter estendere tale interpretazione al caso non lineare. È stata presentata la lettura di un modello non lineare attraverso la varietà riemanniana indotta dalla sua forma matematica e caratterizzata dalla parametrizzazione scelta, immersa in uno spazio ambiente euclideo; sono state introdotte le nozioni classiche note dalla Geometria Differenziale atte allo studio di tali enti matematici, con lo scopo di chiarificare cosa significhi studiare le proprietà di uno spazio, in relazione in particolare alle misure di curvatura. Il lungo, ma necessario, lavoro di richiamo di conoscenze matematiche ha portato infine all'introduzione, nel **quarto capitolo**, delle misure di curvatura di Bates e Watts e della distorsione di Box, basate sull'interpretazione geometrica di cui si è detto; tali strumenti rappresentano un valido supporto per il ricercatore che si trovi di fronte alla scelta di un modello da applicare ai dati in suo possesso. Si è inoltre rimarcato il fatto che l'utilizzo delle misure di curvatura proprie della Geometria Differenziale da parte di questi autori è non sufficientemente esplicitato e in merito a ciò si è provveduto a dimostrare le relazioni intercorrenti omesse da Bates e Watts; a riguardo di una esplicita aderenza all'impiego delle misure di curvature classiche e del calcolo tensoriale in ambito statistico si è citato l'articolo del Prof. Zanella che ne propone un impiego nell'ambito della determinazione delle regioni di confidenza per i parametri dei modelli non lineari.

Le proprietà delle stime parametriche che, come si è dimostrato, sono intimamente legate alle misure di curvatura e di distorsione proposte, possono costituire un importante metro di misura tramite il quale orientare le proprie analisi e valutare la

validità o meno di quelle procedure inferenziali che poggiano su assunzioni di una certa regolarità nella distribuzione delle stime (tipicamente la normalità). Come è noto, nell'ambito dell'analisi di regressione non lineare proprietà desiderabili come la non distorsione e l'efficienza vengono spesso a mancare in larga misura, privando il ricercatore della possibilità di applicare test statistici o altre procedure analoghe con relativa sicurezza. Per quanto concerne questo aspetto è stato visto come l'aumento della dimensione campionaria, la replicazione del disegno sperimentale (quando queste opzioni risultassero attuabili) e la riparametrizzazione possano fornire un valido aiuto nella direzione dell'approssimazione migliore possibile di queste caratteristiche. Queste considerazioni, in principio presentate in via teorica, sono state verificate empiricamente a mezzo di simulazioni su differenti combinazioni modello-data set, fornendo una conferma pratica dell'utilità effettiva che consegue dal tener conto delle misure di curvatura in ambito di regressione non lineare. Operazioni come la riduzione della distorsione nelle stime e la riduzione della variabilità ad esse inerenti hanno un'importanza estrema negli ambiti tecnologici più delicati: ad esempio un modello non lineare può legare le componenti necessarie per la sintesi di un farmaco, ed un errore all'apparenza anche piccolissimo nei dosaggi indicati dalle stime potrebbe causare conseguenze che si possono facilmente intuire catastrofiche. Anche sotto l'aspetto economico una precisione superiore può rivelarsi fondamentale: spesso le prove sperimentali hanno un costo elevatissimo, e ripeterle più volte potrebbe essere impossibile o non conveniente per motivi di natura finanziaria; altresì i parametri di un modello non lineare come quello Weibull presentato nelle esemplificazioni potrebbero costituire indicazioni sulla base delle quali orientare l'acquisto di un nuovo terreno per una coltivazione, il metodo di trattamento del suolo, e così via.

In definitiva l'affascinante branca della Matematica nota come Geometria Differenziale può fornire un importante contributo all'analisi statistica, come strumento accurato e rigoroso per uno studio più approfondito di problematiche complicate di grande interesse applicativo e non solo metodologico, quali sono quelle relative alla modellistica non lineare.

Ringraziamenti

Questa tesi simboleggia la fine di un ciclo di studi, nonché di vita, di cui ho imparato a cogliere ogni giorno di più la crucialità e l'enorme importanza. Questa circostanza così felice richiede prepotentemente che venga data voce alla parte di me che in questi anni è stata formata, aiutata, incoraggiata, ispirata, accompagnata dalle persone che ho vicino nella vita di tutti i giorni, da quelle che ho incontrato e che hanno percorso con me una parte di strada, e da quelle che spero continueranno a percorrerla finché sarà possibile.

Il ringraziamento più profondo e più sentito è diretto senza dubbio alcuno alla mia famiglia. A mio padre Riccardo, che ha passato, e passerà, tutta la mia vita ad insegnarmi con le sue parole ed il suo esempio quanto sia fondamentale improntare alla serietà e all'onestà qualunque cosa verso cui si dirigano i propri sforzi, che mi ha rimproverato duramente in ogni occasione in cui si rendeva conto che questa serietà veniva a mancare da parte mia e che avrei potuto raggiungere risultati di gran lunga migliori se solo avessi messo più impegno nelle cose che stavo facendo, o che semplicemente avrei potuto evitare di combinare qualche sciocchezza. Per aver tentato di insegnarmi la necessità di mettere pazienza e dedizione in ogni cosa, ahimé per questo temo sia necessario ancora un po' di tempo! Per essersi sacrificato tanto nel corso degli anni per me e avermi permesso di arrivare fino a qui, ed in cambio avermi chiesto soltanto di imparare ad ascoltare. Per essere un padre meraviglioso per questi e mille altri motivi, grazie Papà.

A mia madre Elena, il faro nella notte della mia esistenza "sciamannata", come la definisce lei, per avermi cresciuto con tutto l'amore possibile per una madre, e anche di più. Per avermi sempre dato comprensione per i miei errori, per avermi urlato dietro ogni volta che giravo la testa dall'altra parte per non stare a sentire in cosa stessi sbagliando e per essere sempre stata pronta ad accogliermi a braccia aperte nelle rare occasioni in cui la testa la giravo verso di lei. Per aver suscitato in me l'amore per la conoscenza, per avermi insegnato a far nascere un interrogativo di fronte ad ogni cosa che mi si presentasse davanti, per avermi messo un libro in mano ogni volta che le trovava libere da televisori da rompere o persone da tormentare, quand'ero piccolo, per essere riuscita sempre a rappresentare per me

un modello così stupendo di persona. Per essere una enciclopedia con le gambe, per saper preparare i dolci più buoni del mondo e non volerli mangiare perché troppo dolci, per il suo instancabile lavoro di mamma con i suoi figli e per così tanti motivi che non basterebbe un'altra tesi di soli ringraziamenti se non per esporne una minima parte, grazie Mamma.

A mio fratello Emanuele, verso cui ho probabilmente accumulato più colpe che meriti, ma a cui tengo più che alla mia stessa vita, per i pochi ma così felici momenti in cui ci siamo fermati a parlare, per le serate passate gomito a gomito sui monti cuneesi a ridere e fare macello, momenti indimenticabili, per il suo entrare in camera mentre studio e chiedermi "che stai facendo?" e il mostrarsi interessato, salvo poi scuotere sconsolatamente il capo di fronte alla mia folle passione per queste formule così complicate, per tutti i momenti, tristi e felici che abbiamo passato e che passeremo insieme, grazie Loli.

A mia sorella Elisa, un piccolo concentrato d'amore con la grinta di una tigre, che illumina questa casa e rende tutto più bello, per il mio rivedermi in lei alla sua età, per il sorriso sempre stampato sul volto, grazie Bimba!

Ai miei nonni materni, Eros e Leda, due persone eccezionali quali non se ne trovano quasi più su questo pianeta, per avermi coccolato e viziato all'inverosimile in ogni momento che passavo con voi, per avermi amato come se fossi un figlio e per essere in grado di dare questo amore a tutte le persone che vi circondano e farvi da loro ricambiare, per aver contribuito così tanto al fondamentale lavoro di mia madre di farmi crescere nella sete di sapere, nel continuo interrogarsi di fronte alle cose della vita, per aver passato tanto tempo con me a raccontarmi di voi, delle vostre esperienze e delle vostre opinioni, per l'incredibile energia che mettete in ogni cosa che fate, per avermi insegnato l'importanza fondamentale della famiglia. Per essere come siete, sempre e comunque, che piova o ci sia il sole, grazie Papie, grazie Nonna.

A mia nonna paterna, Maria, e alla memoria di nonno Giorgio, per aver cresciuto un uomo come mio padre, e perché nonostante la lontananza e il fatto che abbiamo potuto passare relativamente poco tempo insieme voglio che sappiate che ciò non mi ha impedito di stimare ed ammirare anche voi, grazie di cuore.

A tutti gli altri parenti, vicini e lontani, che in un modo o nell'altro sono stati elementi importanti della mia formazione, grazie.

Un pensiero di affetto infinito e la speranza che la nostra amicizia possa continuare per sempre a Laura, senza la quale probabilmente non sarei già qui a scrivere questi ringraziamenti, per la sua pazienza senza confini nelle tante ore di studio trascorse insieme e nei momenti della mia vita in cui mi sono perso e l'ho fatta soffrire, per essere stata lì, pronta ad accogliermi con un abbraccio quando mi sono reso conto dei miei errori, per avermi fatto riflettere, per avermi incoraggiato, per essere stata sempre così simbiotica con me e per essere la donna che mia madre sogna per me...per avermi aperto gli occhi in tante occasioni e per aver avuto sempre una parola di sdegno per le fanciulle con cui mi vedeva! Grazie Laura, non so come avrei fatto senza di te.

Un ringraziamento importante a Rodolfo, compagno di tante infinite discussioni su argomenti astrusi, per avermi aiutato a coltivare la mia immaginazione e, anche se probabilmente non ne è consapevole, insegnato che non dovevo aver paura di volare con la mente fino a sfiorare anche le cose a prima vista più inaccessibili. Per le lunghe giornate passate a studiare a casa sua, isolati dal mondo e proiettati verso l'infinito, per essere forse l'unica persona che ho conosciuto in vita mia capace di seguirmi nei miei voli pindarici nel mondo della matematica ed esserne altrettanto entusiasta, per le serate devastanti, per i momenti di riflessione, per le tutto ciò che abbiamo condiviso e spero che divideremo ancora, Rodo, PDF!!!!

Ringrazio tutte le persone che ho incontrato e che sono riuscite a farmi vedere le cose sotto punti di vista nuovi ed interessanti, che hanno solleticato la mia immaginazione, che mi hanno apprezzato e che mi hanno spronato. A tale proposito vorrei ringraziare in modo particolare il professor Angelo Zanella, relatore di questa tesi, che senza saperlo ha fatto nascere in me l'amore per una materia che la gente reputa così ostile, come la Statistica, durante il suo corso di Statistica II, e ha contribuito a mantenere questo fuoco acceso, unendo acume e profondità d'analisi alle sue grandi doti di insegnamento, anche nei corsi successivi; forse però il merito più grande lo ha guadagnato durante l'impegnativo lavoro di attenta correzione e revisione di questa tesi che abbiamo svolto insieme, dandomi una fondamentale lezione di professionalità, di atteggiamento di criticità nei confronti del proprio lavoro e di quello degli altri. Ringrazio il professor Diego Zappa, mio relatore per la tesina della triennale, che con il suo stile di insegnamento così vivace e coinvolgente e la

sua personalità così disponibile ed entusiasta ha contribuito non poco a stimolare dentro di me la voglia di continuare per questa strada, mostrandomi più di quel che potessi immaginare quanto la Statistica possa essere emozionante e fonte di gratificazione, lo ringrazio per i frequenti colloqui in cui rubo troppo del suo prezioso tempo ed in cui ascolta le mie considerazioni e mi consiglia come meglio non potrebbe fare.

Ringrazio tutti i professori dell'istituto di Statistica dell'Università Cattolica del Sacro Cuore di Milano, per gli sforzi e l'impegno che giornalmente approfondono nel loro lavoro, e per la simpatia e cordialità che hanno sempre mostrato.

Un grazie è dovuto anche a tutti gli amici, gli amori ed i conoscenti che al di fuori dell'università hanno accompagnato questo mio percorso, condividendo esperienze od interessi, facendomi soffrire, facendomi crescere, facendomi passare delle belle serate, delle splendide vacanze (grazie Ale!), delle campagne indimenticabili (grazie Fra!!), lunghissime stupende nottate al telefono in cui la mattina non arrivava mai (e se non arrivava si scaricava il telefono..eh Silvietta?) e qualunque altra cosa che sia successa, devo molto anche a loro. Grazie anche a Livia per la pazienza e la disponibilità nell'aiutarmi con i miei dubbi matematici, e per avermi posto, probabilmente senza volerlo, davanti ad una riflessione di importanza fondamentale per il mio futuro. Un grazie ad Arrigo, senza il quale non sarei riuscito ad iniziare a studiare questi argomenti così complicati nel modo migliore, la sua chiarezza e la sua disponibilità sono state fondamentali. Ma sto diventando decisamente prolisso, grazie a tutti!

Un ultimo, decisamente egoistico, ringraziamento a me stesso, ma non poteva mancare no? Non sarei più io.. Per aver saputo cogliere davvero qualcosa da ognuno degli infiniti input esterni che mi hanno bombardato nella mia vita e averlo interiorizzato, per avere imparato così tanto di me stesso negli ultimissimi anni ed essere riuscito a ristabilire un equilibrio che pensavo perduto, per essere sempre stato capace di divertirmi con poco ma non averne mai avuto abbastanza di quel poco. Per avere avuto la forza di lanciarmi in questa tesi, che in certi (rari eh!) momenti ho davvero avuto paura costituisse un'impresa più grande di me, e aver tenuto duro quasi senza dir niente a nessuno. Grazie.

Appendice

A.1 La fattorizzazione QR

Data una matrice $\mathbf{A} \in \mathbb{R}^{n \times p}$ ($n \geq p$), si cerca una matrice ortogonale \mathbf{Q} ed una matrice triangolare \mathbf{R} tale che $\mathbf{A} = \mathbf{QR}$. Tale fattorizzazione può essere ottenuta con differenti tecniche, in particolare mediante l'utilizzo delle matrici elementari di Householder o di Givens, oppure tramite il metodo di ortogonalizzazione di Gram-Schmidt.

Si considerino i vettori colonna della matrice \mathbf{A}

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_p]$$

e siano essi linearmente indipendenti (ovvero la matrice abbia rango p). Applicando alla sequenza $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p$ la procedura di ortogonalizzazione di Gram-Schmidt otteniamo:

$$\begin{aligned} \mathbf{f}_1 &= \mathbf{a}_1 \\ \mathbf{f}_2 &= \mathbf{a}_2 - \frac{\langle \mathbf{a}_2, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2} \mathbf{f}_1 \\ \mathbf{f}_3 &= \mathbf{a}_3 - \frac{\langle \mathbf{a}_3, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2} \mathbf{f}_1 - \frac{\langle \mathbf{a}_3, \mathbf{f}_2 \rangle}{\|\mathbf{f}_2\|^2} \mathbf{f}_2 \\ &\dots \\ \mathbf{f}_p &= \mathbf{a}_p - \frac{\langle \mathbf{a}_p, \mathbf{f}_1 \rangle}{\|\mathbf{f}_1\|^2} \mathbf{f}_1 - \dots - \frac{\langle \mathbf{a}_p, \mathbf{f}_{p-1} \rangle}{\|\mathbf{f}_{p-1}\|^2} \mathbf{f}_{p-1} \end{aligned}$$

dove con $\langle -, - \rangle$ si è indicato l'usuale prodotto scalare fra vettori. La sequenza

$$\left\{ \mathbf{q}_1 = \frac{\mathbf{f}_1}{\|\mathbf{f}_1\|}, \mathbf{q}_2 = \frac{\mathbf{f}_2}{\|\mathbf{f}_2\|}, \dots, \mathbf{q}_p = \frac{\mathbf{f}_p}{\|\mathbf{f}_p\|} \right\}$$

è una base ortonormale dello spazio generato dalle colonne della matrice \mathbf{A} . Si considerino ora le due matrici

$$\mathbf{Q} = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_p]$$

$$\mathbf{R} = \begin{bmatrix} \|\mathbf{f}_1\| & \langle \mathbf{a}_2, \mathbf{q}_1 \rangle & \langle \mathbf{a}_3, \mathbf{q}_1 \rangle & \dots & \langle \mathbf{a}_p, \mathbf{q}_1 \rangle \\ 0 & \|\mathbf{f}_2\| & \langle \mathbf{a}_3, \mathbf{q}_2 \rangle & \dots & \langle \mathbf{a}_p, \mathbf{q}_2 \rangle \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \|\mathbf{f}_p\| \end{bmatrix}$$

Si verifica che $\mathbf{A} = \mathbf{QR}$. Una tale fattorizzazione è detta *QR non normalizzata*. Viceversa, viene detta *scomposizione QR normalizzata* la fattorizzazione *QR* dove \mathbf{Q} è una matrice ortogonale $n \times n$ ed \mathbf{R} una matrice $n \times p$ tale per cui

$$\mathbf{R} = \begin{pmatrix} \tilde{\mathbf{R}} \\ \mathbf{0} \end{pmatrix}$$

dove $\tilde{\mathbf{R}}$ è una matrice triangolare alta $p \times p$ e $\mathbf{0}$ è una matrice $(n-p) \times p$ di zeri. La fattorizzazione non normalizzata può essere ottenuta, come visto, tramite la procedura di ortogonalizzazione di Gram-Schmidt, quella normalizzata invece richiede l'impiego di successive moltiplicazioni di matrici di Householder (o di Givens), si veda a riguardo R. A. Horn e C. R. Johnson, *Matrix Analysis*, Sezione 2.6, Cambridge University Press, 1985. In questo modo la matrice \mathbf{Q} contiene sia una base per lo spazio generato dalle colonne di \mathbf{A} che una base per il suo complemento ortogonale, ovvero per lo spazio perpendicolare a tali colonne. La fattorizzazione *QR* di una matrice è unica in virtù del fatto che lo è la matrice \mathbf{Q} (per la dimostrazione si rimanda ancora al testo di Horn e Johnson).

A.2 Coincidenza della misura di Bates e Watts con la curvatura normale

La seguente definizione è tratta da Geometria 2, Sernesi (2001), p.292.

Sia $\mathbf{e} \in T_{\mathbf{x}}(S)$ un versore tangente alla superficie S nel punto \mathbf{x} . La curvatura normale di S in \mathbf{x} nella direzione \mathbf{e} è:

$$\kappa(\mathbf{e}) := \langle -\nabla_{\mathbf{e}} \mathbf{n}, \mathbf{e} \rangle = \langle W_p(\mathbf{e}), \mathbf{e} \rangle = II(\mathbf{e}, \mathbf{e}).$$

Dove \mathbf{n} è la normale alla superficie in \mathbf{x} , W_p è l'operatore di Weingarten e $II(-, -)$ la Seconda Forma Fondamentale.

Si consideri ora l'espressione della curvatura proposta da Bates e Watts

$$K_{\mathbf{h}}^N = \frac{\|\ddot{\eta}_{\mathbf{h}}^N\|}{\|\dot{\eta}_{\mathbf{h}}\|^2}$$

Esplicitiamo numeratore e denominatore con riferimento alle impostazioni del problema, così come presentato nel paragrafo 4.3. Facciamo uso della convenzione sulla somma degli indici ripetuti, i quali verranno indicati con lettere greche, come d'uso nella notazione tensoriale; tali lettere non indicheranno quindi mai un elevamento a potenza, anche se poste ad apice. Qualora necessari gli elevamenti a potenza saranno sempre indicati da una parentesi intorno alla base della potenza. Chiameremo inoltre in questa sede per comodità

$$\eta_{\alpha} = \frac{\partial \eta_{\mathbf{h}}}{\partial \theta^{\alpha}}$$

$$\eta_{\alpha\beta} = \frac{\partial \eta_{\mathbf{h}}}{\partial \theta^{\alpha} \partial \theta^{\beta}}, \quad \alpha, \beta = 1, 2, \dots, p$$

in questo caso in particolare $p = 2$.

Possiamo quindi scomporre il vettore curvatura come segue:

$$\begin{aligned}\ddot{\eta}_{\mathbf{h}} &= \frac{\partial \dot{\eta}_{\mathbf{h}}}{\partial s} = \frac{\partial}{\partial s} \left(\eta_{\alpha} \frac{d\theta^{\alpha}}{ds} \right) = \frac{\partial}{\partial s} (\eta_{\alpha}) \frac{d\theta^{\alpha}}{ds} + \eta_{\alpha} \frac{d^2\theta^{\alpha}}{ds^2} = \\ &= \left(\frac{\partial \eta_{\alpha}}{\partial s \theta^{\beta}} \frac{d\theta^{\beta}}{ds} \right) \frac{d\theta^{\alpha}}{ds} + \eta_{\alpha} \frac{d^2\theta^{\alpha}}{ds^2} = \\ &= \eta_{\alpha\beta} \frac{d\theta^{\alpha}}{ds} \frac{d\theta^{\beta}}{ds} + \eta_{\alpha} \frac{d^2\theta^{\alpha}}{ds^2}\end{aligned}$$

Osserviamo che la quantità $\|\ddot{\eta}_{\mathbf{h}}^{\mathbf{N}}\|$ individua il modulo della proiezione delle vettore delle derivate seconde (il vettore curvatura) sulla normale, ovvero $\|\ddot{\eta}_{\mathbf{h}}^{\mathbf{N}}\| = \langle \ddot{\eta}_{\mathbf{h}}, \mathbf{n} \rangle$.

Di conseguenza possiamo sostituire:

$$\begin{aligned}\|\ddot{\eta}_{\mathbf{h}}^{\mathbf{N}}\| &= \langle \ddot{\eta}_{\mathbf{h}}, \mathbf{n} \rangle = \left\langle \eta_{\alpha\beta} \frac{d\theta^{\alpha}}{ds} \frac{d\theta^{\beta}}{ds} + \eta_{\alpha} \frac{d^2\theta^{\alpha}}{ds^2}, \mathbf{n} \right\rangle = \\ &= \left\langle \eta_{\alpha\beta} \frac{d\theta^{\alpha}}{ds} \frac{d\theta^{\beta}}{ds}, \mathbf{n} \right\rangle + \left\langle \eta_{\alpha} \frac{d^2\theta^{\alpha}}{ds^2}, \mathbf{n} \right\rangle\end{aligned}$$

ma $\left\langle \eta_{\alpha} \frac{d^2\theta^{\alpha}}{ds^2}, \mathbf{n} \right\rangle = 0$ per costruzione, in quanto η_{α} è perpendicolare ad \mathbf{n} , quindi

$$\begin{aligned}\langle \ddot{\eta}_{\mathbf{h}}, \mathbf{n} \rangle &= \left\langle \eta_{\alpha\beta} \frac{d\theta^{\alpha}}{ds} \frac{d\theta^{\beta}}{ds}, \mathbf{n} \right\rangle + \left\langle \eta_{\alpha} \frac{d^2\theta^{\alpha}}{ds^2}, \mathbf{n} \right\rangle = \\ &= \left\langle \eta_{\alpha\beta} \frac{d\theta^{\alpha}}{ds} \frac{d\theta^{\beta}}{ds}, \mathbf{n} \right\rangle = \langle \eta_{\alpha\beta}, \mathbf{n} \rangle \frac{d\theta^{\alpha}}{ds} \frac{d\theta^{\beta}}{ds}\end{aligned}$$

ora, ricordando le definizioni dei coefficienti di primo e secondo tensore fondamentale:

$$\begin{aligned}g_{\alpha\beta} &= \langle \eta_{\alpha}, \eta_{\beta} \rangle \\ b_{\alpha\beta} &= \langle \eta_{\alpha\beta}, \mathbf{n} \rangle\end{aligned}$$

osserviamo che vale la seguente uguaglianza:

$$\begin{aligned} \langle \eta_{\alpha\beta}, \mathbf{n} \rangle \frac{d\theta^\alpha}{ds} \frac{d\theta^\beta}{ds} &= \frac{\langle \eta_{\alpha\beta}, \mathbf{n} \rangle d\theta^\alpha d\theta^\beta}{(ds)^2} = \frac{(d\varpi)^2}{(ds)^2} = \\ &= \frac{b_{\alpha\beta} d\theta^\alpha d\theta^\beta}{g_{\alpha\beta} d\theta^\alpha d\theta^\beta} = \frac{II(\mathbf{h}, \mathbf{h})}{I(\mathbf{h}, \mathbf{h})} \end{aligned}$$

Se scegliamo \mathbf{h} di norma unitaria allora $I(\mathbf{h}, \mathbf{h})=1$, e questo è sempre possibile, equivale a parametrizzare per lunghezza d'arco la curva tagliata dal piano normale alla superficie, vale

$$\|\dot{\mathbf{i}}_{\mathbf{h}}^N\| = \langle \ddot{\mathbf{i}}_{\mathbf{h}}, \mathbf{n} \rangle = \langle \eta_{\alpha\beta}, \mathbf{n} \rangle \frac{d\theta^\alpha}{ds} \frac{d\theta^\beta}{ds} = II(\mathbf{h}, \mathbf{h}) = \kappa(\mathbf{h})$$

che è la curvatura normale del luogo delle soluzioni lungo la direzione \mathbf{h} . Inoltre il denominatore dell'espressione è $\|\dot{\mathbf{i}}_{\mathbf{h}}\|^2$, che nell'ipotesi di prendere un versore (quindi con la curva parametrizzata per lunghezza d'arco) vale esattamente 1, come risulta evidente sviluppandolo:

$$\begin{aligned} \|\dot{\mathbf{i}}_{\mathbf{h}}\|^2 &= \left(\eta_\alpha \frac{d\theta^\alpha}{ds} \right)^2 = \left(\eta_1 \frac{d\theta^1}{ds} \right)^2 + 2\eta_1\eta_2 \frac{d\theta^1}{ds} \frac{d\theta^2}{ds} + \left(\eta_2 \frac{d\theta^2}{ds} \right)^2 = \\ &= g_{\alpha\beta} \frac{d\theta^\alpha}{ds} \frac{d\theta^\beta}{ds} = \frac{g_{\alpha\beta} d\theta^\alpha d\theta^\beta}{(ds)^2} = \frac{g_{\alpha\beta} d\theta^\alpha d\theta^\beta}{g_{\alpha\beta} d\theta^\alpha d\theta^\beta} = 1 \end{aligned}$$

Quindi

$$K_{\mathbf{h}}^N = \frac{\|\dot{\mathbf{i}}_{\mathbf{h}}^N\|}{\|\dot{\mathbf{i}}_{\mathbf{h}}\|^2} = II(\mathbf{h}, \mathbf{h}) = \kappa(\mathbf{h})$$

questo ragionamento è valido sempre, perché è sempre possibile considerare la curva lungo la direzione \mathbf{h} e giacente nel piano normale intersecante la superficie

come parametrizzata per lunghezza unitaria. Nel nostro caso in particolare abbiamo che:

$$\begin{aligned}
K_{\mathbf{h}}^N &= \frac{\|\dot{\eta}_{\mathbf{h}}^N\|}{\|\dot{\eta}_{\mathbf{h}}\|^2} = \frac{\langle \ddot{\eta}_{\mathbf{h}}, \mathbf{n} \rangle}{\|\dot{\eta}_{\mathbf{h}}\|^2} = \frac{\left\langle \sum_{\alpha, \beta=1}^2 \frac{\partial^2 \eta}{\partial \theta^\alpha \partial \theta^\beta} h^\alpha h^\beta, \mathbf{n} \right\rangle}{\sum_{t=1}^3 \left(\sum_{\alpha=1}^2 \frac{\partial \eta_t}{\partial \theta^\alpha} h^\alpha \right)^2} = \\
&= \frac{\sum_{\alpha, \beta=1}^2 \left\langle \frac{\partial^2 \eta}{\partial \theta^\alpha \partial \theta^\beta}, \mathbf{n} \right\rangle h^\alpha h^\beta}{(h^1)^2 \sum_{t=1}^3 \left(\frac{\partial \eta_t}{\partial \theta^1} \right)^2 + 2h^1 h^2 \sum_{t=1}^3 \frac{\partial \eta_t}{\partial \theta^1} \frac{\partial \eta_t}{\partial \theta^2} + (h^2)^2 \sum_{t=1}^3 \left(\frac{\partial \eta_t}{\partial \theta^2} \right)^2} = \\
&= \frac{b_{\alpha\beta} h^\alpha h^\beta}{g_{\alpha\beta} h^\alpha h^\beta}
\end{aligned}$$

e parametrizzando la curva sulla superficie nella direzione \mathbf{h} per la lunghezza d'arco otteniamo che $g_{\alpha\beta} h^\alpha h^\beta = 1$, da cui segue $K_{\mathbf{h}}^N \equiv \kappa(\mathbf{h})$.

Il fatto che tale misura risulti invariante rispetto a cambiamenti del sistema di coordinate sulla superficie del modello (ovvero una riparametrizzazione) può essere semplicemente dimostrato tramite l'ausilio delle leggi di trasformazione tensoriale.

Consideriamo una generica trasformazione di coordinate $x \rightarrow y$ tale da rispettare le condizioni di invertibilità e differenziabilità (determinante dello Jacobiano della trasformazione non nullo), ovvero tale per cui valgono le relazioni:

$$\begin{cases} y^1 = y^1(x^1, x^2, \dots, x^p) \\ y^2 = y^2(x^1, x^2, \dots, x^p) \\ \vdots \\ y^p = y^p(x^1, x^2, \dots, x^p) \end{cases} \Leftrightarrow \begin{cases} x^1 = x^1(y^1, y^2, \dots, y^p) \\ x^2 = x^2(y^1, y^2, \dots, y^p) \\ \vdots \\ x^p = x^p(y^1, y^2, \dots, y^p) \end{cases}$$

Ancora una volta verrà sfruttata la convenzione sulla sommatoria degli indici ripetuti.

Sono note (si veda a riguardo un qualunque testo di calcolo tensoriale) le seguenti regole di trasformazione:

Regola di trasformazione di un vettore controvariante

$$dy^\alpha = \frac{\partial y^\alpha}{\partial x^\beta} dx^\beta$$

Regola di trasformazione di un tensore doppio covariante

$$A'_{\alpha\beta} = \frac{\partial x^\gamma}{\partial y^\alpha} \frac{\partial x^\delta}{\partial y^\beta} A_{\gamma\delta}$$

Primo e Secondo Tensore Fondamentale sono interamente covarianti di ordine 2, seguono quindi le seguenti relazioni notevoli:

Invarianza dello scalare $(ds)^2$ rispetto a cambiamenti di coordinate sulla superficie

$$(ds)_y^2 = \bar{g}_{\alpha\beta} dy^\alpha dy^\beta = g_{\gamma\delta} \frac{\partial x^\gamma}{\partial y^\alpha} \frac{\partial x^\delta}{\partial y^\beta} \frac{\partial y^\alpha}{\partial x^\gamma} dx^\gamma \frac{\partial y^\beta}{\partial x^\delta} dx^\delta = g_{\gamma\delta} dx^\gamma dx^\delta = (ds)_x^2$$

Invarianza dello scalare $(d\omega)^2$ rispetto a cambiamenti di coordinate sulla superficie

$$(d\omega)_y^2 = \bar{b}_{\alpha\beta} dy^\alpha dy^\beta = b_{\gamma\delta} \frac{\partial x^\gamma}{\partial y^\alpha} \frac{\partial x^\delta}{\partial y^\beta} \frac{\partial y^\alpha}{\partial x^\gamma} dx^\gamma \frac{\partial y^\beta}{\partial x^\delta} dx^\delta = b_{\gamma\delta} dx^\gamma dx^\delta = (d\omega)_x^2$$

Dove con il pedice x o y si è indicato che si sta lavorando rispettivamente nelle parametrizzazioni x o y .

Sia detta Σ la superficie di interesse (nel nostro caso sarà il luogo delle soluzioni della stima parametrica secondo i minimi quadrati, ovvero la superficie η) e $T_p(\Sigma)$ il piano tangente nel punto p a tale superficie (nel nostro caso sarà il punto di stima θ_0).

Siano ora $w \in T_p(\Sigma), \|w\|=1, \bar{w}, \|\bar{w}\|=1$ la sua rappresentazione rispetto alle coordinate y e $k_{n,*}(\bullet)$ la curvatura normale nel punto p della superficie lungo la direzione \bullet e nella parametrizzazione $*$. Vale:

$$k_{n,x}(w) = II_x(w, w) = b_{\gamma\delta} dx^\gamma dx^\delta = (d\varpi)_x^2 = (d\varpi)_y^2 = \bar{b}_{\alpha\beta} dy^\alpha dy^\beta = II_y(\bar{w}, \bar{w}) = k_{n,y}(\bar{w})$$

Segue quindi l'asserto, ovvero che la misura di curvatura intrinseca di Bates e Watts è invariante rispetto a cambiamenti di parametrizzazione del modello, essendo coincidente con la curvatura normale di una superficie lungo una fissata direzione, che si è appena dimostrato essere invariante dalla parametrizzazione della superficie stessa.

A.3 Listati del codice impiegato per le simulazioni

```
#Modello E(Y)=X^theta

>x <- c(2,3)
>y <- c(2.5,10)
>datimod1 <- data.frame(y,x)
>fitmod1 <- nls(y~x^theta,datimod1,start=c(theta=2))

Nonlinear regression model
  model: y ~ x^theta
  data:  datimod1
  theta
2.053690
  residual sum-of-squares: 2.933361

Parameters:
      Estimate Std. Error t value Pr(>|t|)
theta  2.0537    0.1575   13.04  0.0487 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.713 on 1 degrees of freedom
#-----
#Modello E(Y)=X^log(phi)

>fitmod2 <- nls(y~x^log(phi),datimod1,start=c(phi=5))

Nonlinear regression model
  model: y ~ x^log(phi)
  data:  datimod1
  phi
7.796609
  residual sum-of-squares: 2.933361

Parameters:
      Estimate Std. Error t value Pr(>|t|)
phi    7.797    1.228    6.35  0.0994 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.713 on 1 degrees of freedom
#-----
#Simulazioni theta e phi

>sim1 <- function(nns) {

  x1 <- c(2,3)
  y1 <- c()
  thetaSim <- c()
  phiSim <- c()
  tvartheta <- 0.024799
  tvarphi <- 1.507470

  skew <- function(x){

    n <- length(x)
    s3 <- sqrt(var(x)*(n-1)/n)^3
    mx <- mean(x)
    sk <- sum((x-mx)^3)/s3

    sk/n

  }

}
```

```

kurt <- function(x){

    n <- length(x)
    s4 <- sqrt(var(x)*(n-1)/n)^4
    mx <- mean(x)
    kt <- sum((x-mx)^4)/s4

    kt/n - 3

}

for (i in 1:nns) {

    y1[1] <- x1[1]^summary(fitmod1)$parameters[1,1] +
    rnorm(1,0,summary(fitmod1)$sigma)
    y1[2] <- x1[2]^summary(fitmod1)$parameters[1,1] +
    rnorm(1,0,summary(fitmod1)$sigma)
    datiTempS <- data.frame(y1,x1)
    thetaSim[i] <-
summary(nls(y1~x1^thetaS,datiTempS,start=c(thetaS=2)))$parameters[1,1]
    phiSim <- exp(thetaSim)

}

thetaSimNorm <- (thetaSim-mean(thetaSim))/sd(thetaSim)
phiSimNorm <- (phiSim-mean(phiSim))/sd(phiSim)

par(mfcol=c(2,2))

hist(thetaSimNorm,freq=F, xlab="theta", ylab="", main="", col="red")
curve(dnorm(x),add=TRUE,col="blue")
qqnorm(thetaSimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(thetaSimNorm,col="blue")

hist(phiSimNorm,freq=F, xlab="phi", ylab="", main="", col="red")
curve(dnorm(x),add=TRUE,col="blue")
qqnorm(phiSimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(phiSimNorm,col="blue")

par(mfcol=c(1,1))

print("Parametro theta")
print(c("Skewness      :",skew(thetaSimNorm)))
print(c("Excess Kurtosis  :",kurt(thetaSimNorm)))
print(c("% Bias          :",100*(mean(thetaSim)-
summary(fitmod1)$parameters[1,1])/summary(fitmod1)$parameters[1,1]))
print(c("% Excess Variance :",100*(sd(thetaSim)^2-tvartheta)/tvartheta))
print("Parametro phi")
print(c("Skewness      :",skew(phiSimNorm)))
print(c("Excess Kurtosis  :",kurt(phiSimNorm)))
print(c("% Bias          :",100*(mean(phiSim)-
summary(fitmod2)$parameters[1,1])/summary(fitmod2)$parameters[1,1]))
print(c("% Excess Variance :",100*(sd(phiSim)^2-tvarphi)/tvarphi))

}
> sim1(2000)
[1] "Parametro theta"
[1] "Skewness      : " "-0.599682547178424"
[1] "Excess Kurtosis  : " "0.717474632628631"
[1] "% Bias          : " "-0.701225175444717"
[1] "% Excess Variance : " "14.9647393920032"
[1] "Parametro phi"
[1] "Skewness      : " "-0.0524177592874428"
[1] "Excess Kurtosis  : " "0.0474334038619899"
[1] "% Bias          : " "-0.0609385010091507"
[1] "% Excess Variance : " "5.82805196761389"

```



```

#Modello di Michaelis-Menten, p=2 n=3

>concl <- c(0.02,0.22,1.1)
>rate1 <- c(47,159,200)
>datimod2n3 <- data.frame(rate1,concl)
>mmjh <- deriv3(~ Vm * concl/(K + concl), c("Vm", "K"), function(Vm, K, concl) NULL)

>fitmod2 <- nls(rate1 ~ mmjh(Vm, K, concl), data = datimod2n3, start = list(Vm=200,
K=0.1))

>plot(rate1 ~ concl, data = datimod2n3, las = 1,
      xlab = "Concentrazione del Substrato (ppm)",
      ylab = "Velocità di Reazione (counts/min/min)",
      main = "Modello di Michaelis Menten adattato ai dati con n=3")

>xxx <- seq(0, 1.2, len = 100)
>lines(xxx, predict(fitmod2, list(concl=xxx)), lty = 1, col = 2)

>MMmod1 <- function(a,b) a*concl/(b+concl)

>graf.mod <- function(modello, parsmin, parsmax, prec, angl, ang2, bb, MM) {

      a <- seq(parsmin[1], parsmax[1], length=prec)
      b <- seq(parsmin[2], parsmax[2], length=prec)

      zz <- outer(a,b,modello)

      persp(zz, col="green", theta=ang1, phi=ang2, box=bb, main=MM)

}

>graf.mod(MMmod1, c(190,0), c(200,10), 36, 145, 15, T, "Luogo delle Soluzioni ")

>simlmm <- function(nns) {

  x1 <- concl
  theta12Sim <- matrix(, nns, 2)
  theta1 <- summary(fitmod2)$parameters[1,1]
  theta2 <- summary(fitmod2)$parameters[2,1]

  tvartheta1 <- summary(fitmod2)$cov.unscaled[1,1]
  tvartheta2 <- summary(fitmod2)$cov.unscaled[2,2]

  skew <- function(x){

    n <- length(x)
    s3 <- sqrt(var(x)*(n-1)/n)^3
    mx <- mean(x)
    sk <- sum((x-mx)^3)/s3

    sk/n

  }

  kurt <- function(x){

    n <- length(x)
    s4 <- sqrt(var(x)*(n-1)/n)^4
    mx <- mean(x)
    kt <- sum((x-mx)^4)/s4

    kt/n - 3

  }

}

```

```

for (i in 1:nns) {

  y1 <- theta1*x1/(theta2+x1) + rnorm(3,0,summary(fitmod2)$sigma)
  datiTempS <- data.frame(y1,x1)
  theta12Sim[i,1] <-
summary(nls(y1~th1*x1/(th2+x1),datiTempS,start=c(th1=212.4864,th2=0.07216885),algori
thm="port"))$parameters[1,1]
  theta12Sim[i,2] <-
summary(nls(y1~th1*x1/(th2+x1),datiTempS,start=c(th1=212.4864,th2=0.07216885),algori
thm="port"))$parameters[2,1]

}

theta1SimNorm <- (theta12Sim[,1]-mean(theta12Sim[,1]))/sd(theta12Sim[,1])
theta2SimNorm <- (theta12Sim[,2]-mean(theta12Sim[,2]))/sd(theta12Sim[,2])

par(mfcol=c(2,2))

hist(theta1SimNorm,freq=F, xlab="theta1", ylab="", main="", col="red")
curve(dnorm(x),add=TRUE,col="blue")
qqnorm(theta1SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(theta1SimNorm,col="blue")

hist(theta2SimNorm,freq=F, xlab="theta2", ylab="", main="", col="red")
curve(dnorm(x),add=TRUE,col="blue")
qqnorm(theta2SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(theta2SimNorm,col="blue")

par(mfcol=c(1,1))

print("Parametro theta1")
print(c("Skewness      :",skew(theta1SimNorm)))
print(c("Excess Kurtosis  :",kurt(theta1SimNorm)))
print(c("% Bias          :",100*(mean(theta12Sim[,1])-theta1)/theta1))
print(c("% Excess Variance :",100*(sd(theta12Sim[,1])^2-tvartheta1)/tvartheta1))

print("Parametro theta2")
print(c("Skewness      :",skew(theta2SimNorm)))
print(c("Excess Kurtosis  :",kurt(theta2SimNorm)))
print(c("% Bias          :",100*(mean(theta12Sim[,2])-theta2)/theta2))
print(c("% Excess Variance :",100*(sd(theta12Sim[,2])^2-tvartheta2)/tvartheta2))

}

>box.bias1 <- function(expr,pars,dti) { #Si creino un oggetto expr (il modello)
#e uno pars (i nomi dei parametri)

  DD <- function(expr,pars, order = 1) {

    if(order < 1) stop("'order' must be >= 1")
    if(order == 1) D(expr,pars)
    else DD(D(expr, pars), pars, order - 1)

  }

  param <- c()

  for (uu in 1:length(pars)) {

    param[uu] <- summary(fitmod2)$parameters[uu,1]

  }

  th1 <- param[1]

```

```

th2 <- param[2]

x <- dti$concl
matr <- matrix(, length(x), length(param))
harray <- array(, dim=c(length(param), length(param), length(x)))

for (nn in 1:length(param)) {
    matr[,nn] <- eval(DD(expr, pars[nn], order=1))
}

FF <- c()

for (i in 1:length(x)) {
    FF[i] <- crossprod(matr[i,])
}

SFF <- 1/(sum(FF))

for (ff in 1:length(param)) {
    for (jj in 1:length(param)) {
        harray[ff, jj, ] <- eval(DD(DD(expr, pars[ff]), pars[jj]))
    }
}

sumdh <- matrix(, length(x), length(pars))

for (j in 1:length(x)) {
    sumdh[j, ] <- matr[j, ]*sum(diag(harray[, , j]*SFF))
}

v.somme <- c()

for (mm in 1:length(pars)) {
    v.somme[mm] <- sum(sumdh[, mm])
}

bias <- c()

for (pp in 1:length(pars)) {
    bias[pp] <- -((summary(fitmod2)$sigma^2)/2)*SFF*v.somme[pp]
}

p.bias <- (bias/param)*100

print("Il bias percentuale e' :")
print(t(p.bias))
}

```

```

#-----
#Modello di Michaelis-Menten, p=2 n=6

>conc2 <- c(0.02,0.06,0.11,0.22,0.56,1.1)
>rate2 <- c(76,107,123,159,201,200)
>datimod2n6 <- data.frame(rate2,conc2)
>mmjh2 <- deriv3(~ Vm * conc2/(K + conc2), c("Vm", "K"), function(Vm, K, conc2)
NULL)

>fitmod2 <- nls(rate2 ~ mmjh2(Vm, K, conc2), data = datimod2n6, start = list(Vm=200,
K=0.1))

>plot(rate2 ~ conc2, data = datimod2n6, las = 1,
      xlab = "Concentrazione del Substrato (ppm)",
      ylab = "Velocità di Reazione (counts/min/min)",
      main = "Modello di Michaelis Menten adattato ai dati con n=6")

>xxx <- seq(0, 1.2, len = 100)
>lines(xxx, predict(fitmod2, list(conc2=xxx)), lty = 1, col = 2)

Nonlinear regression model
  model: rate2 ~ mmjh2(Vm, K, conc2)
  data: datimod2n6
           Vm           K
202.81063312  0.04665259
residual sum-of-squares: 404.6183

Parameters:
      Estimate Std. Error t value Pr(>|t|)
Vm 2.028e+02  8.206e+00  24.715 1.59e-05 ***
K  4.665e-02  8.243e-03   5.659 0.00480 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.06 on 4 degrees of freedom

[1] "Il bias percentuale e':"
      [,1]      [,2]
[1,] -7.402195e-08 0.5926951

Parameter effects: c^theta x sqrt(F) = 0.3585
                   c^iota  x sqrt(F) = 0.1651

>simlmm <- function(nns) {

  x1 <- conc2
  theta12Sim <- matrix(,nns,2)
  theta1 <- summary(fitmod2)$parameters[1,1]
  theta2 <- summary(fitmod2)$parameters[2,1]

  tvartheta1 <- summary(fitmod2)$cov.unscaled[1,1]
  tvartheta2 <- summary(fitmod2)$cov.unscaled[2,2]

  skew <- function(x){

    n <- length(x)
    s3 <- sqrt(var(x)*(n-1)/n)^3
    mx <- mean(x)
    sk <- sum((x-mx)^3)/s3

    sk/n

  }

  kurt <- function(x){

```

```

        n <- length(x)
        s4 <- sqrt(var(x)*(n-1)/n)^4
        mx <- mean(x)
        kt <- sum((x-mx)^4)/s4

        kt/n - 3
    }

    for (i in 1:nns) {

        y1 <- theta1*x1/(theta2+x1) + rnorm(length(x1),0,summary(fitmod2)$sigma)
        datiTempS <- data.frame(y1,x1)
        theta12Sim[i,1] <-
summary(nls(y1~th1*x1/(th2+x1),datiTempS,start=c(th1=212.4864,th2=0.07216885),algori
thm="port"))$parameters[1,1]
        theta12Sim[i,2] <-
summary(nls(y1~th1*x1/(th2+x1),datiTempS,start=c(th1=212.4864,th2=0.07216885),algori
thm="port"))$parameters[2,1]

    }

    theta1SimNorm <- (theta12Sim[,1]-mean(theta12Sim[,1]))/sd(theta12Sim[,1])
    theta2SimNorm <- (theta12Sim[,2]-mean(theta12Sim[,2]))/sd(theta12Sim[,2])

    par(mfcol=c(2,2))

    hist(theta1SimNorm,freq=F, xlab="theta1", ylab="", main="", col="red")
    curve(dnorm(x),add=TRUE,col="blue")
    qqnorm(theta1SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
    qqline(theta1SimNorm,col="blue")

    hist(theta2SimNorm,freq=F, xlab="theta2", ylab="", main="", col="red")
    curve(dnorm(x),add=TRUE,col="blue")
    qqnorm(theta2SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
    qqline(theta2SimNorm,col="blue")

    par(mfcol=c(1,1))

    print("Parametro theta1")
    print(c("Skewness      :",skew(theta1SimNorm)))
    print(c("Excess Kurtosis  :",kurt(theta1SimNorm)))
    print(c("% Bias          :",100*(mean(theta12Sim[,1])-theta1)/theta1))
    print(c("% Excess Variance :",100*(sd(theta12Sim[,1])^2-tvartheta1)/tvartheta1))

    print("Parametro theta2")
    print(c("Skewness      :",skew(theta2SimNorm)))
    print(c("Excess Kurtosis  :",kurt(theta2SimNorm)))
    print(c("% Bias          :",100*(mean(theta12Sim[,2])-theta2)/theta2))
    print(c("% Excess Variance :",100*(sd(theta12Sim[,2])^2-tvartheta2)/tvartheta2))

}

```

```

#-----
#Modello di Michaelis-Menten, p=2 n=12, replicazione del piano completo

>conc3 <- c(0.02,0.02,0.06,0.06,0.11,0.11,0.22,0.22,0.56,0.56,1.1,1.1)
>rate3 <- c(76,47,97,107,123,139,159,152,191,201,207,200)
>datimod2n12r <- data.frame(rate3,conc3)
>mmjh3 <- deriv3(~ Vm * conc3/(K + conc3), c("Vm", "K"), function(Vm, K, conc3)
NULL)

>fitmod2 <- nls(rate3 ~ mmjh3(Vm, K, conc3), data = datimod2n12r, start =
list(Vm=200, K=0.1))

>plot(rate3 ~ conc3, data = datimod2n6r, las = 1,
      xlab = "Concentrazione del Substrato (ppm)",
      ylab = "Velocità di Reazione (counts/min/min)",
      main = "Modello di Michaelis Menten adattato ai dati con n=12\n Replicazione
del Caso 2.")

>xxx <- seq(0, 1.2, len = 100)
>lines(xxx, predict(fitmod2, list(conc3=xxx)), lty = 1, col = 2)

Nonlinear regression model
model: rate3 ~ mmjh3(Vm, K, conc3)
data: datimod2n12r
      Vm      K
212.6836298  0.0641211
residual sum-of-squares: 1195.449

Parameters:
      Estimate Std. Error t value Pr(>|t|)
Vm 2.127e+02  6.947e+00  30.615 3.24e-11 ***
K  6.412e-02  8.281e-03   7.743 1.57e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.93 on 10 degrees of freedom

Parameter effects: c^theta x sqrt(F) = 0.2121
                   c^iota  x sqrt(F) = 0.092

#-----
#Modello di Michaelis-Menten, p=2 n=12, replicazione del piano completo

>conc3 <- c(0.02,0.02,0.06,0.06,0.11,0.11,0.22,0.22,0.56,0.56,1.1,1.1)
>rate3 <- c(76,47,97,107,123,139,159,152,191,201,207,200)
>datimod2n12r <- data.frame(rate3,conc3)
>mmjh3 <- deriv3(~ Vm* conc3/(K + conc3), c("Vm", "K"), function(Vm, K, conc3) NULL)
>fitmod2 <- nls(rate3 ~ mmjh3(Vm, K, conc3), data = datimod2n12r, start =
list(Vm=200, K=0.01))

>plot(rate3 ~ conc3, data = datimod2n6r, las = 1,
      xlab = "Concentrazione del Substrato (ppm)",
      ylab = "Velocità di Reazione (counts/min/min)",
      main = "Modello di Michaelis Menten adattato ai dati con n=12\n Replicazione
del Caso 2.")
>xxx <- seq(0, 1.2, len = 100)
>lines(xxx, predict(fitmod2, list(conc3=xxx)), lty = 1, col = 2)

>simlmm <- function(nns) {

  x1 <- conc3
  theta12Sim <- matrix(,nns,2)
  theta1 <- summary(fitmod2)$parameters[1,1]
  theta2 <- summary(fitmod2)$parameters[2,1]

  tvartheta1 <- summary(fitmod2)$cov.unscaled[1,1]
  tvartheta2 <- summary(fitmod2)$cov.unscaled[2,2]

```

```

skew <- function(x){
  n <- length(x)
  s3 <- sqrt(var(x)*(n-1)/n)^3
  mx <- mean(x)
  sk <- sum((x-mx)^3)/s3

  sk/n
}

kurt <- function(x){
  n <- length(x)
  s4 <- sqrt(var(x)*(n-1)/n)^4
  mx <- mean(x)
  kt <- sum((x-mx)^4)/s4

  kt/n - 3
}

for (i in 1:nns) {
  y1 <- theta1*x1/(theta2+x1) + rnorm(length(x1),0,summary(fitmod2)$sigma)
  datiTempS <- data.frame(y1,x1)
  theta12Sim[i,1] <-
summary(nls(y1~th1*x1/(th2+x1),datiTempS,start=c(th1=212.4864,th2=0.07216885)))$parameters[1,1]
  theta12Sim[i,2] <-
summary(nls(y1~th1*x1/(th2+x1),datiTempS,start=c(th1=212.4864,th2=0.07216885)))$parameters[2,1]
}

theta1SimNorm <- (theta12Sim[,1]-mean(theta12Sim[,1]))/sd(theta12Sim[,1])
theta2SimNorm <- (theta12Sim[,2]-mean(theta12Sim[,2]))/sd(theta12Sim[,2])

par(mfcol=c(2,2))

hist(theta1SimNorm,freq=F, xlab="theta1", ylab="", main="", col="red")
curve(dnorm(x),add=TRUE,col="blue")
qqnorm(theta1SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(theta1SimNorm,col="blue")

hist(theta2SimNorm,freq=F, xlab="theta2", ylab="", main="", col="red")
curve(dnorm(x),add=TRUE,col="blue")
qqnorm(theta2SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(theta2SimNorm,col="blue")

par(mfcol=c(1,1))

print("Parametro theta1")
print(c("Skewness      :",skew(theta1SimNorm)))
print(c("Excess Kurtosis  :",kurt(theta1SimNorm)))
print(c("% Bias          :",100*(mean(theta12Sim[,1])-theta1)/theta1))
print(c("% Excess Variance :",100*(sd(theta12Sim[,1])^2-tvartheta1)/tvartheta1))

print("Parametro theta2")
print(c("Skewness      :",skew(theta2SimNorm)))
print(c("Excess Kurtosis  :",kurt(theta2SimNorm)))
print(c("% Bias          :",100*(mean(theta12Sim[,2])-theta2)/theta2))
print(c("% Excess Variance :",100*(sd(theta12Sim[,2])^2-tvartheta2)/tvartheta2))
}

```

```

#Onion Bulbs Plus Tops, Yield vs Time, Modello Weibull

>ox <- c(1:15)
>oy <-
c(16.08,33.83,65.80,97.20,191.55,326.20,386.87,520.53,590.03,651.92,724.93,699.56,68
9.96,637.56,717.41)
>odata <- data.frame(ox,oy)

>weib1 <- deriv3(~ Theta1-Theta2*exp(-Theta3*(ox^Theta4)),
c("Theta1","Theta2","Theta3","Theta4"), function(Theta1,Theta2,Theta3,Theta4,ox)
NULL)

>fitweib <- nls(oy ~ weib1(Theta1,Theta2,Theta3,Theta4,ox), data = odata, start =
list(Theta1=690,Theta2=673,Theta3=0.001,Theta4=3))

>plot(oy ~ ox, data = odata, las = 1,
      xlab = "Tempo",
      ylab = "Resa",
      main = "Modello di tipo Weibull")

>xxx <- seq(0, 16, len = 100)
>lines(xxx, predict(fitweib, list(ox=xxx)), lty = 1, col = 2)

Nonlinear regression model
  model: oy ~ weib1(Theta1, Theta2, Theta3, Theta4, ox)
  data: odata
      Theta1      Theta2      Theta3      Theta4
6.950401e+02 6.735109e+02 1.522441e-03 3.261564e+00
  residual sum-of-squares: 7834.297

Parameters:
      Estimate Std. Error t value Pr(>|t|)
Theta1 6.950e+02  1.317e+01  52.793 1.39e-14 ***
Theta2 6.735e+02  2.398e+01  28.091 1.36e-11 ***
Theta3 1.522e-03  1.017e-03   1.496  0.163
Theta4 3.262e+00  3.344e-01   9.753 9.48e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.69 on 11 degrees of freedom

>box.bias <- function(expr,pars,dti) {      #Si creino un oggetto expr (il modello)
                                           #e uno pars (i nomi dei parametri)

  DD <- function(expr,pars, order = 1) {

    if(order < 1) stop("'order' must be >= 1")
    if(order == 1) D(expr,pars)
    else DD(D(expr, pars), pars, order - 1)

  }

  param <- c()

  for (uu in 1:length(pars)) {

    param[uu] <- summary(fitweibr)$parameters[uu,1]

  }

  Th1 <- param[1]
  Th2 <- param[2]
  GMA <- param[3]
  Th4 <- param[4]

```



```

x <- dti$ox
matr <- matrix(,length(x),length(param))
harray <- array(,dim=c(length(param),length(param),length(x)))

for (nn in 1:length(param)) {
  matr[,nn] <- eval(DD(expr,pars[nn],order=1))
}

FF <- c()
for (i in 1:length(x)) {
  FF[i] <- crossprod(matr[i,])
}

SFF <- 1/(sum(FF))

for (ff in 1:length(param)) {
  for (jj in 1:length(param)) {
    harray[ff,jj,] <- eval(DD(DD(expr,pars[ff]),pars[jj]))
  }
}

sumdh <- matrix(,length(x),length(pars))
for (j in 1:length(x)) {
  sumdh[j,] <- matr[j,]*sum(diag(harray[, ,j]*SFF))
}

v.somme <- c()
for (mm in 1:length(pars)) {
  v.somme[mm] <- sum(sumdh[,mm])
}

bias <- c()
for (pp in 1:length(pars)) {
  bias[pp] <- -((summary(fitweib)$sigma^2)/2)*SFF*v.somme[pp]
}

p.bias <- (bias/param)*100

print("Il bias percentuale e' :")
print(t(p.bias))
}

>simlweibr <- function(nns) {
  x1 <- ox
  theta12Sim <- matrix(,nns,4)
  theta1 <- summary(fitweibr)$parameters[1,1]
  theta2 <- summary(fitweibr)$parameters[2,1]

```

```

GAMMA <- summary(fitweibr)$parameters[3,1]
theta4 <- summary(fitweibr)$parameters[4,1]

tvartheta1 <- summary(fitweibr)$cov.unscaled[1,1]
tvartheta2 <- summary(fitweibr)$cov.unscaled[2,2]
tvarGAMMA <- summary(fitweibr)$cov.unscaled[3,3]
tvartheta4 <- summary(fitweibr)$cov.unscaled[4,4]

skew <- function(x){
  n <- length(x)
  s3 <- sqrt(var(x)*(n-1)/n)^3
  mx <- mean(x)
  sk <- sum((x-mx)^3)/s3

  sk/n
}

kurt <- function(x){
  n <- length(x)
  s4 <- sqrt(var(x)*(n-1)/n)^4
  mx <- mean(x)
  kt <- sum((x-mx)^4)/s4

  kt/n - 3
}

for (i in 1:nns) {
  y1 <- theta1-theta2*exp(-exp(-GAMMA)*x1^theta4) +
  rnorm(length(x1),0,summary(fitweibr)$sigma)
  datiTempS <- data.frame(y1,x1)
  tempfit <- nls(y1~th1-th2*exp(-exp(-
th3)*x1^th4),datiTempS,start=c(th1=6.950401e+02,th2=6.735109e+02,th3=6.487439,th4=3.
261564),algorithm="port")
  theta2Sim[i,1] <- summary(tempfit)$parameters[1,1]
  theta2Sim[i,2] <- summary(tempfit)$parameters[2,1]
  theta2Sim[i,3] <- summary(tempfit)$parameters[3,1]
  theta2Sim[i,4] <- summary(tempfit)$parameters[4,1]
}

theta1SimNorm <- (theta2Sim[,1]-mean(theta2Sim[,1]))/sd(theta2Sim[,1])
theta2SimNorm <- (theta2Sim[,2]-mean(theta2Sim[,2]))/sd(theta2Sim[,2])
theta3SimNorm <- (theta2Sim[,3]-mean(theta2Sim[,3]))/sd(theta2Sim[,3])
theta4SimNorm <- (theta2Sim[,4]-mean(theta2Sim[,4]))/sd(theta2Sim[,4])

par(mfcol=c(2,4))

hist(theta1SimNorm,freq=F, xlab="theta1", ylab="", main="", col="red")
curve(dnorm(x),add=TRUE,col="blue")
qqnorm(theta1SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(theta1SimNorm,col="blue")

hist(theta2SimNorm,freq=F, xlab="theta2", ylab="", main="", col="red")
curve(dnorm(x),add=TRUE,col="blue")
qqnorm(theta2SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(theta2SimNorm,col="blue")

hist(theta3SimNorm,freq=F, xlab="gamma", ylab="", main="", col="red")
curve(dnorm(x),add=TRUE,col="blue")

```

```

qqnorm(theta3SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(theta3SimNorm,col="blue")

hist(theta4SimNorm,freq=F,xlab="theta4",ylab="",main="",col="red")
curve(dnorm(x),add=TRUE,col="blue")
qqnorm(theta4SimNorm,main="",xlab="Quantili Teorici",ylab="Quantili
Campionari",col="red")
qqline(theta4SimNorm,col="blue")

par(mfcol=c(1,1))

print("Parametro theta1")
print(c("Skewness      :",skew(theta1SimNorm)))
print(c("Excess Kurtosis  :",kurt(theta1SimNorm)))
print(c("% Bias          :",100*(mean(theta12Sim[,1])-theta1)/theta1))
print(c("% Excess Variance :",100*(sd(theta12Sim[,1])^2-tvartheta1)/tvartheta1))

print("Parametro theta2")
print(c("Skewness      :",skew(theta2SimNorm)))
print(c("Excess Kurtosis  :",kurt(theta2SimNorm)))
print(c("% Bias          :",100*(mean(theta12Sim[,2])-theta2)/theta2))
print(c("% Excess Variance :",100*(sd(theta12Sim[,2])^2-tvartheta2)/tvartheta2))

print("Parametro gamma")
print(c("Skewness      :",skew(theta3SimNorm)))
print(c("Excess Kurtosis  :",kurt(theta3SimNorm)))
print(c("% Bias          :",100*(mean(theta12Sim[,3])-GAMMA)/GAMMA))
print(c("% Excess Variance :",100*(sd(theta12Sim[,3])^2-tvarGAMMA)/tvarGAMMA))

print("Parametro theta4")
print(c("Skewness      :",skew(theta4SimNorm)))
print(c("Excess Kurtosis  :",kurt(theta4SimNorm)))
print(c("% Bias          :",100*(mean(theta12Sim[,4])-theta4)/theta4))
print(c("% Excess Variance :",100*(sd(theta12Sim[,4])^2-tvartheta4)/tvartheta4))

}

#-----
#Modello Weibull Riparametrizzato

>weiblr <- deriv3(~ Theta1-Theta2*exp(-exp(-GAMMA)*(ox^Theta4)),
c("Theta1","Theta2","GAMMA","Theta4"), function(Theta1,Theta2,GAMMA,Theta4,ox) NULL)

>fitweibr <- nls(oy ~ weiblr(Theta1,Theta2,GAMMA,Theta4,ox), data = odata, start =
list(Theta1=690,Theta2=673,GAMMA=2,Theta4=3))

```

Bibliografia

- BATES, D.M. e WATTS, D.G. (1980). Relative Curvature Measures of Nonlinearity. *J. R. Statistic. Soc., Ser. B* 42, 1, No. 1-25.
- BARNDORFF-NIELSEN, O. E., COX, D. R. e REID, N. (1986). The Role of Differential Geometry in Statistical Theory. *Int. Stat. Rev.* 54, 83-96.
- BOX, M. J. (1971). Bias in Nonlinear Estimation. *J. R. Statistic. Soc., Ser. B* 33, No. 2, 171-201.
- FINZI, B. e PASTORI, M. (1961). *Calcolo Tensoriale ed Applicazioni*. Zanichelli Editore, Bologna.
- IACUS, S. M. e MASAROTTO, G. (2003). *Laboratorio di Statistica con R*. Milano, McGraw-Hill.
- RATKOWSKY, D.A. (1983). *Nonlinear Regression Modeling. A unified Practical Approach*. Dekker, New York.
- SEBER, G. A. J. e WILD, C. J. (1989). *Nonlinear Regression*. Wiley, New York.
- SERNESI, E. (2001). *Geometria 2*. Bollati Boringhieri, Torino.
- ZANELLA, A. (1997). *Lezioni di Teoria dei Campioni. Parte I: Elementi di teoria della programmazione ed analisi statistica degli esperimenti*. Istituto di Statistica, Università Cattolica del S.Cuore di Milano.
- ZANELLA, A. (1974). Sulle Regioni di Confidenza per i Parametri dei Modelli Non Lineari. I. *Calcolo*, 3, 365-401.
- ZANELLA, A. (1975). Sulle Regioni di Confidenza per i Parametri dei Modelli Non Lineari. II. *Calcolo*, 1, 1-37.

Materiale online

Dispense di Geometria Differenziale e di Calcolo Differenziale Assoluto (Calcolo Tensoriale) ad opera di Arrigo Amadori disponibili all'indirizzo

<http://www.arrigoamadori.com/lezioni/Matematica.htm>

Introduction to Tensor Calculus and Continuum Mechanics, di HEINBOCKEL, J. H., Dipartimento di Matematica e Statistica, Old Dominion University.
<http://www.math.odu.edu/~jhh/counter2.html>

Intrinsic Geometry of Surfaces, di ALEKSANDROV, A. D. e ZALGALLER, V. A.
http://www.ams.org/online_bks/mmono15/

Course of Differential Geometry, di SHARIPOV, R. A.
<http://www.geocities.com/r-sharipov/r4-b3.htm>

Why to Calculate, When to Use, and How to Understand Curvature Measures of Nonlinearity, KAROLCZAK, M.
<http://www.currentseparations.com/issues/14-1/cs14-1b.pdf>