



Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche

Dottorato di Ricerca in

Metodologia Statistica per la Ricerca Scientifica

XX ciclo

**Esperimenti per
Modelli Parzialmente Lineari con
Applicazione ai Computer Experiments**

Maroussa Zagoraiou

Dipartimento di Scienze Statistiche "P. Fortunati"

Marzo 2008



Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche

Dottorato di Ricerca in

Metodologia Statistica per la Ricerca Scientifica

XX ciclo

Esperimenti per
Modelli Parzialmente Lineari con
Applicazione ai Computer Experiments

Maroussa Zagoraiou

Tutor

Prof.ssa Alessandra Giovagnoli

Coordinatore

Prof.ssa Daniela Cocchi

Settore Disciplinare

SECS-S/01

Dipartimento di Scienze Statistiche "P. Fortunati"

Marzo 2008

A mio padre... μου λείπεις

Ringraziamenti

Sono grata alla Prof.ssa Alessandra Giovagnoli che, con la sua grande preparazione e professionalità, è stata per me una guida verso un traguardo reso possibile grazie soprattutto alla sua costante presenza e disponibilità e che ha fatto sì che dai momenti più difficili traessi spunto e motivazione per andare avanti.

I lavori di alcune persone sono resi unici non solo dall'impegno protratto in essi e dalle capacità individuali, ma soprattutto dalla voglia di arrivare in fondo, non perché esso sia dovuto ma perché è desiderato, dettato da una spinta verso l'alto, verso il raggiungimento di una meta. E' anche grazie al Dott. Alessandro Baldi Antognini che riesco a credere al connubio tra scienza e passione.

Ringrazio i miei colleghi di dottorato e dello studio di via Zamboni, in particolare Marta e Michele, figure importanti lungo questo cammino. A loro due va la mia gratitudine per essere stati punti di forza nel perseguire i miei risultati che non prescindevano mai dalla loro presenza. Colleghi nello studio, amici nella vita. Inoltre, ritengo indispensabile ringraziare Valentina, che con il suo grande appoggio è stata per me punto di riferimento in un periodo fondamentale di questo dottorato.

Un caloroso ringraziamento va a Efi e Laskarò che sono sempre state capaci di far sentire la loro presenza malgrado la lontananza fisica, a Patrizia per la sua profonda amicizia e a Giulia che ha dovuto sopportarmi in casa. Un pensiero particolare va a Felice, insostituibile figura nella mia vita.

Questa tesi di dottorato è dedicata a una parte di me che non c'è più.

Maroussa Zagoraïou

Indice

	5
1 Pianificazione degli Esperimenti per Modelli Parametrici	11
1.1 Modelli Lineari, non Lineari e Parzialmente Lineari	11
1.2 Piani Sperimentali per Modelli Lineari e non	14
1.2.1 Problemi Lineari	14
1.2.2 Problemi non Lineari	18
1.3 Piani Sequenziali	21
1.3.1 Procedura Sequenziale di Chaudhuri e Mykland	21
2 Modelli per Computer Experiments	25
2.1 Esperimenti Simulati	25
2.2 Metamodelli	26
2.2.1 Obiettivo del Metamodello	26
2.2.2 Procedura di Sviluppo del Metamodello	27
2.3 Modelli basati sulle Funzioni-base	28
2.3.1 Modelli Polinomiali	30
2.3.2 Splines	30
2.3.3 Fourier	31
2.4 Random Field Regression Models	32
2.4.1 Modelli di Correlazione	33
2.4.2 Previsori	33
2.4.3 Metodi Stocastici per Predizioni Spaziali (Kriging)	38
3 Esperimenti per la Stima del Metamodello: Disegni Space-Filling	41
3.1 Premessa	41

3.2	Disegni basati su Metodi di Campionamento	42
3.2.1	Campionamento Casuale Semplice e Stratificato	42
3.2.2	Campionamento da Ipercubi Latini	43
3.3	Disegni basati sulla Distanza	45
3.4	Disegni Uniformi	46
3.5	Disegni basati su Combinazioni di Criteri	47
3.6	Applicazione della Metodologia Kriging e dei Disegni Space-Filling	48
3.6.1	Introduzione	48
3.6.2	Metodologia	49
4	Disegni Ottimi per il Kriging Ordinario	53
4.1	Premessa	53
4.2	Modello	55
4.3	La Matrice di Informazione di Fisher	56
4.4	I Disegni Ottimi per le Stime Separate di μ e θ	61
4.5	Il Disegno della Progressione Geometrica	64
4.5.1	Determinazione della Dimensione Campionaria	66
4.5.2	Efficienza del Disegno della Progressione Geometrica	69
5	Disegni Sequenziali per il Modello di Decadimento Esponenziale	71
5.1	Il Modello di Decadimento Esponenziale	71
5.2	Procedure Sequenziali per il Modello di Decadimento Esponenziale	73
5.2.1	Procedura basata sulla Massima Verosimiglianza	74
5.2.2	Procedura basata sulla Parziale Linearità	75
5.3	Confronti	76
5.4	Conclusioni	78
A	La Maggiorizzazione	79

Elenco delle figure

2.1	Illustrazione dell'impatto del trend nel caso di interpolazione.	39
2.2	Illustrazione dell'impatto del trend nel caso di estrapolazione.	40
3.1	Grafico dell'errore quadratico medio di x_1 e x_2	51
3.2	Plot di iterazione tra x_1 e x_2 basato sul metamodello previsivo del quarto step.	52
4.1	$I_\mu(\xi_n(r))$ in funzione di n	67
4.2	$I_\theta(\xi_n(0.3))$ in funzione di n per $\theta = 0.5$ (sinistra), $\theta = 1$ (centrale) e $\theta = 5$ (destra).	69
4.3	Grafici dell'efficienza riguardante μ (blu) e θ (rosa) del GPD $\xi_{15}(r)$ con $\theta = 0.5$ (sinistra), $\theta = 2$ (centrale) e $\theta = 5$ (destra).	70
4.4	Grafici dell'efficienza riguardante θ del GPD $\xi_{15}(r)$ al variare di r	70
4.5	Grafici dell'efficienza riguardante μ del GPD $\xi_{15}(r)$ al variare di r	70

Introduzione

Il metodo sperimentale è caposaldo per l'acquisizione di conoscenze in numerosi campi scientifici. La teoria classica della pianificazione di esperimenti e l'intuizione che la sperimentazione vada eseguita secondo opportune regole è legata soprattutto al nome di Sir Ronald A. Fisher con il suo libro sul disegno degli esperimenti (Fisher, 1935). L'interesse di Fisher era rivolto agli esperimenti agricoli quando era responsabile per la statistica e l'analisi dei dati alla Rothamsted Agricultural Experimental Station. Successivamente si è sviluppata la teoria dei disegni ottimi, associata al nome di Jack Kiefer, il quale nel 1959 presentò le sue idee innovative alla Royal Statistical Society, a Londra.

Vi è una vasta letteratura sui disegni ottimi nella teoria della pianificazione di esperimenti per la stima dei parametri di modelli lineari. Un campo, invece, ancora poco esplorato è quello della pianificazione di esperimenti per modelli non lineari in cui i problemi di ottimalità del piano sono di difficile soluzione. La complicazione fondamentale inerente a questo tipo di modelli è la dipendenza della soluzione del problema di ottimalità dai parametri sconosciuti presenti nel modello, il che spesso dà luogo a disegni la cui ottimalità dipende dalla conoscenza dei parametri stessi (ottimalità locale).

Nella mia tesi mi sono occupata di alcuni aspetti particolari del problema in questione, iniziando il mio studio presso il Department of Statistics della London School of Economics and Political Science, dove sono attivi alcuni esperti della materia, tra cui il Prof. Henry P. Wynn e il Prof. Antony C. Atkinson, che desidero ringraziare. Ho proseguito la mia attività di ricerca presso il Dipartimento di Scienze Statistiche "Paolo Fortunati", sotto la guida della Prof.ssa Alessandra Giovagnoli.

Struttura della tesi

L'oggetto d'interesse di questa tesi sono i modelli parzialmente lineari. Definiamo così modelli tali che il valore atteso della risposta è lineare rispetto a un sottoinsieme dei parametri presenti nel modello. La tesi si articola essenzialmente in tre parti.

La prima parte è sostanzialmente introduttiva e riguarda la presentazione delle idee fondamentali inerenti alla pianificazione ottima di esperimenti per modelli lineari, parzialmente lineari e non lineari.

La seconda parte, costituita da tre capitoli, tratta gli esperimenti al computer. Negli ultimi decenni si assiste con tendenza crescente alla sostituzione di esperimenti fisici con esperimenti virtuali: all'osservazione di un fenomeno nella realtà viene sostituito l'output di un simulatore, cioè un modello matematico implementato al computer che descrive nel modo più fedele possibile la dipendenza di una variabile di risposta da un insieme di fattori di input. Questo avviene sia a causa dell'elevato dispendio di risorse temporali impiegate da un esperimento reale, sia perché i codici, o simulatori, diventano uno strumento di uso sempre più economico. Tuttavia, dovendo descrivere in modo accurato un sistema fisico, un codice di simulazione tende ad essere per sua stessa natura molto complesso e pertanto è stato suggerito di ricorrere ai cosiddetti metamodelli, cioè a modelli molto più semplici che ne costituiscono una valida approssimazione. Il metamodello è spesso di tipo parametrico e si ricorre ad un esperimento con poche prove del simulatore (training set) per la stima dei parametri.

In particolare, nel secondo capitolo viene presentata una breve rassegna dei modelli dei computer experiments approfondendo la metodologia Kriging che rappresenta una delle tecniche più diffuse per modellare l'output di un computer experiment. Due sono gli approcci solitamente utilizzati per la scelta del disegno sperimentale in questo contesto, allo scopo di stimare i parametri presenti nei metamodelli parametrici. Il primo si basa sulla teoria dei disegni ottimi; il secondo prevede l'utilizzo dei cosiddetti disegni Space-Filling, ossia procedure che sono spesso basate su metodi di campionamento e mirano a distribuire i punti del dominio in modo equispaziato motivate dall'esigenza di osservare la risposta lungo l'intero spazio di disegno.

L'oggetto del terzo capitolo concerne i disegni Space-Filling. Verrà anche esposta un'applicazione riguardante il campo della telefonia mobile tratta dalla letteratura dove gli autori identificano il settaggio ideale, ossia la scelta dei valori delle variabili di ingresso che ottimizzano la risposta, attraverso la metodologia Kriging abbinata ad una procedura sequenziale.

L'argomento del quarto capitolo è la ricerca di piani di esperimento per la stima di massima verosimiglianza dei parametri del Kriging Ordinario unidimensionale avente struttura di correlazione esponenziale parametrizzata. Il problema consiste nello scegliere il disegno,

ossia l'insieme di punti di osservazione, che minimizza la matrice di varianza covarianza degli stimatori di massima verosimiglianza dei parametri incogniti del modello la quale può essere approssimata attraverso l'inversa della matrice di informazione di Fisher. I principali risultati da me raggiunti possono essere così riassunti: esaminando singolarmente la componente di trend μ ed il parametro di correlazione θ , è stato dimostrato che il disegno ottimo per la stima del trend è quello equispaziato, ovvero il disegno composto da punti fra loro equidistanziati, mentre per quanto riguarda la stima di θ il piano sperimentale ideale tende a sovrapporre tutti i punti di osservazione agli estremi del dominio. E' evidente l'esistenza di un palese contrasto dal punto di vista della precisione di stima dei due parametri, in quanto il disegno ottimo per μ rappresenta la scelta pessima per θ e viceversa. Alla luce del comportamento conflittuale in questione, nella tesi è stata quindi proposta una nuova classe di procedure che consentono di ottenere un opportuno compromesso in termini di precisione di stima simultanea di entrambi i parametri, garantendo al tempo stesso alta efficienza singolarmente per ciascuno di essi. Vengono illustrate notevoli proprietà riguardanti la classe di cui sopra sia per via analitica che per via grafica.

Nel quinto capitolo, ovvero nella terza parte della tesi, si considera un diverso modello parzialmente lineare, ossia il modello di decadimento esponenziale che trova le sue applicazioni principalmente nel campo della farmacocinetica la quale ha come obiettivo lo studio dell'evoluzione temporale della concentrazione di un farmaco nell'organismo. Vengono formulate due proposte di esperimento per la stima dei suoi parametri attraverso delle procedure sequenziali che comprendono l'adozione del piano di campionamento delle unità statistiche e la scelta di una regola di determinazione delle assegnazioni. Le proprietà di queste procedure vengono studiate per mezzo di una serie di simulazioni effettuate con l'utilizzo del software Matlab. Questo capitolo si conclude con delle brevi considerazioni legate ai possibili sviluppi futuri.

Capitolo 1

Pianificazione degli Esperimenti per Modelli Parametrici

1.1 Modelli Lineari, non Lineari e Parzialmente Lineari

L'obiettivo della modellistica è quello di trovare un modello in grado di rappresentare fenomeni di qualunque natura. Il modello è una costruzione ipotetica di un sistema, cioè di una porzione della realtà, che contiene elementi fra loro interrelati. Per la costruzione di esso si individuano delle variabili esplicative o indipendenti $\mathbf{x} = (x_1, \dots, x_s)^t$, le cui variazioni provocano la modifica del livello di un'altra grandezza, detta variabile di risposta o dipendente Y , e si cerca di derivare una legge che descrive in maniera sufficientemente accurata il fenomeno di interesse.

Da un punto di vista analitico, possiamo avere vari tipi di modelli. I modelli deterministici sono i più semplici: in essi, la variabile di risposta assume lo stesso valore a fronte degli stessi valori delle variabili esplicative

$$Y = f(\mathbf{x}) = f(x_1, \dots, x_s).$$

Al contrario, i modelli stocastici o probabilistici forniscono risultati in termini di probabilità e assumono la seguente forma:

$$Y \sim F_Y(y | \mathbf{x}) = F_Y(y | x_1, \dots, x_s),$$

dove $F_Y(\cdot)$ è una distribuzione di probabilità. In genere i modelli stocastici hanno una struttura più complessa di quelli deterministici ma sono anche più affidabili in quanto, tenendo conto del caso, sono capaci di fornire risultati più aderenti alla realtà.

Accade di frequente che per descrivere il comportamento della variabile di risposta Y in un modello stocastico, ci si limita a specificare il suo valore atteso $E(Y|\mathbf{x})$ e la sua varianza $V(Y|\mathbf{x})$. Inoltre, molto spesso si assume che il fenomeno possa essere descritto tramite un *modello parametrico* in cui la distribuzione della variabile di risposta dipende da una o più variabili indipendenti \mathbf{x} attraverso una determinata funzione nota a meno di un numero finito di parametri $\boldsymbol{\beta}$.

Il *modello lineare* assume la seguente forma:

$$E(Y|\mathbf{x}) = \sum_{i=1}^p f_i(\mathbf{x})\beta_i \quad V(Y|\mathbf{x}) = \sigma^2, \quad (1.1)$$

in cui σ^2 è una costante che non dipende da \mathbf{x} . Nella (1.1) \mathbf{x} è un vettore $s \times 1$ di variabili esplicative i cui elementi rappresentano combinazioni di livelli di s fattori e β_i è la i -esima componente di un vettore $\boldsymbol{\beta}$, $p \times 1$, di parametri sconosciuti. Le f_i rappresentano particolari trasformazioni delle variabili. Formalmente:

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_p(\mathbf{x}))^t.$$

I *modelli lineari generalizzati* (McCullagh e Nelder, 1989) sono appunto una generalizzazione dei modelli lineari e vanno nominati a parte a causa della loro importanza. I modelli di questo tipo vengono caratterizzati dal fatto che la risposta attesa dipende, per mezzo di una funzione $h(\cdot)$ nota, monotona e differenziabile, da una combinazione lineare dei parametri sconosciuti, il cosiddetto *predittore lineare*:

$$E(Y|\mathbf{x}) = h(\mathbf{f}(\mathbf{x})^t \boldsymbol{\beta}),$$

essendo \mathbf{x} il vettore delle variabili esplicative. Inoltre nei modelli lineari generalizzati si suppone che la distribuzione di probabilità delle osservazioni sia un membro della famiglia esponenziale, ad esempio normale, Poisson, gamma o binomiale. In virtù di ciò, per questi modelli la varianza delle osservazioni può non essere costante.

I *modelli di regressione non lineari* hanno la seguente forma:

$$E(Y|\mathbf{x}) = \tau(\mathbf{x}, \boldsymbol{\beta}) \quad V(Y|\mathbf{x}) = \sigma^2(\mathbf{x}, \boldsymbol{\beta}), \quad (1.2)$$

in cui τ e σ^2 sono funzioni qualunque che dipendono dal vettore di variabili esplicative \mathbf{x} e dai parametri $\boldsymbol{\beta}$. Nel *modello di regressione non lineare classico* si assume che Y sia distribuito normalmente e σ^2 sia costante.

Diamo ora una definizione di *modello parzialmente lineare*:

Definizione 1.1 Sia Y una variabile di risposta la cui distribuzione dipende da \mathbf{x} e dai parametri vettoriali $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Chiameremo un modello *parzialmente lineare* (PLM: Partially Linear Model) se, per \mathbf{x} fissato, esso soddisfa la seguente proprietà

$$\frac{\partial}{\partial \boldsymbol{\beta}_1} E(Y) = \phi(\boldsymbol{\beta}_2, \mathbf{x}),$$

ossia se il valore atteso della risposta è lineare rispetto a un sottoinsieme dei parametri presenti nel modello e la varianza della risposta è una funzione qualunque che dipende dal vettore di variabili esplicative e dai parametri.

E' facile vedere che i modelli lineari possono essere visti come caso particolare di modelli parzialmente lineari, mentre i modelli lineari generalizzati non rientrano nella definizione 1.1. Un'altra definizione data in letteratura è quella di Hill (1980), ovvero

Definizione 1.2 Si consideri il modello di regressione non lineare con Y una variabile di risposta la cui distribuzione dipende da \mathbf{x} e da $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ e σ^2 costante. Hill definisce tale modello *parzialmente non lineare* se

$$\frac{\partial}{\partial \boldsymbol{\beta}} E(Y) = \mathbf{A}(\boldsymbol{\beta}_1) \phi(\boldsymbol{\beta}_2, \mathbf{x}),$$

dove $\mathbf{A}(\boldsymbol{\beta}_1)$ è una matrice quadrata, non singolare che non dipende da \mathbf{x} , mentre $\phi(\boldsymbol{\beta}_2, \mathbf{x})$ è un vettore di funzioni che dipende da \mathbf{x} e da un sottoinsieme, $\boldsymbol{\beta}_2$, dei parametri nel modello.

Si noti che la nostra definizione 1.1 si discosta da quella data da Hill come mostra il seguente esempio:

Esempio 1.3 *Si consideri il seguente modello*

$$E(Y|\mathbf{x}) = \beta_{11} \cos(\beta_2 \mathbf{x}) + \beta_{12} \sin(\beta_2 \mathbf{x}), \quad V(Y|\mathbf{x}) = \sigma^2. \quad (1.3)$$

Indicando con $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})^t$, il modello (1.3) è parzialmente lineare nel senso della definizione 1.1 dato che

$$\frac{\partial}{\partial \boldsymbol{\beta}_1} E(Y) = \begin{pmatrix} \frac{\partial}{\partial \beta_{11}} E(Y) \\ \frac{\partial}{\partial \beta_{12}} E(Y) \end{pmatrix} = \begin{pmatrix} \cos(\beta_2 \mathbf{x}) \\ \sin(\beta_2 \mathbf{x}) \end{pmatrix}$$

risulta funzione di β_2 e di \mathbf{x} e non dipende da $\boldsymbol{\beta}_1$, mentre non rientra nella definizione di Hill. Infatti

$$\frac{\partial}{\partial \boldsymbol{\beta}} E(Y) = \begin{pmatrix} \cos(\beta_2 \mathbf{x}) \\ \sin(\beta_2 \mathbf{x}) \\ \mathbf{x} (\beta_{12} \cos(\beta_2 \mathbf{x}) - \beta_{11} \sin(\beta_2 \mathbf{x})) \end{pmatrix}$$

non può essere espresso come prodotto di una matrice quadrata, non singolare dipendente solo da β_1 e un vettore che dipende solamente da \mathbf{x} e β_2 .

1.2 Piani Sperimentali per Modelli Lineari e non

1.2.1 Problemi Lineari

La teoria classica dei piani di esperimento si preoccupa della stimabilità delle quantità di interesse e della possibilità di eseguire l'analisi della varianza; si cercano piani di esperimento con alcune proprietà che sono intuitivamente attraenti e spesso facilitano l'analisi dei dati. La costruzione di piani con queste proprietà ha portato alla considerazione di svariati problemi di matematica combinatoria il cui studio continua ai giorni nostri. Nel 1959 Kiefer presentò le sue idee innovative attraverso un approccio diverso: una teoria approssimata dei piani di esperimento, valida per grandi campioni. Questa è nota oggi come la teoria dei *disegni ottimi*. Lo scopo di questo capitolo è di introdurre le idee fondamentali di Kiefer e i suoi sviluppi per vari modelli parametrici. La teoria classica dei piani di esperimento assieme alla teoria dei disegni ottimi hanno trovato una sistemazione in eccellenti libri come quello di Atkinson e Donev (1992).

Supponiamo di voler raccogliere un certo numero di osservazioni allo scopo di valutare l'effetto di alcuni fattori su una variabile di risposta Y . In un esperimento i fattori, ossia le variabili esplicative, sono in qualche modo sotto il controllo dello sperimentatore, che entro certi limiti può scegliere di variarne a piacere le modalità (livelli). Ipotizzato un modello lineare adeguato, esistono svariati piani sperimentali tra cui scegliere. Effettuando n osservazioni indipendenti e non correlate e indicando con \mathbf{X} la matrice $n \times p$, il cui elemento generico i, j è dato da $f_j(\mathbf{x}_i)$ e con \mathbf{Y} il vettore la cui i -esima componente è Y_i , il modello (1.1) può essere riscritto

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} \quad V(\mathbf{Y}) = \sigma^2 \mathbf{I}_n, \quad (1.4)$$

in cui \mathbf{I}_n rappresenta la matrice identità. Se il rango di \mathbf{X} è p , $\boldsymbol{\beta}$ può essere stimato

mediante il metodo dei minimi quadrati, cioè:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y}$$

per cui vale

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad V(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^t \mathbf{X})^{-1},$$

con la ben nota proprietà fornita dal Teorema di Gauss-Markov, ossia $\hat{\boldsymbol{\beta}}$ è lo stimatore più efficiente nella classe degli stimatori lineari e non distorti.

Sia Ξ l'insieme dei disegni possibili, dunque

$$\Xi = \{\xi = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}, \mathbf{x}_i \in \mathcal{X}\}, \quad (1.5)$$

in cui n è un numero naturale fisso e \mathcal{X} è il cosiddetto *spazio di disegno*. Avendo stabilito che lo scopo dell'esperimento è ottenere stime ottime di $\boldsymbol{\beta}$ o una predizione ottima, si può fare il confronto tra disegni considerando la matrice di varianza covarianza degli stimatori dei minimi quadrati di $\boldsymbol{\beta}$, ottenuta effettuando un disegno ξ , o equivalentemente la sua inversa. La matrice

$$\mathbf{X}^t \mathbf{X} = \sum_i \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^t, \quad (1.6)$$

che coincide a meno del fattore di proporzionalità σ^{-2} con l'inversa della matrice di varianza covarianza di $\hat{\boldsymbol{\beta}}$, è definita matrice di informazione del disegno ξ . Un esperimento sarà migliore di un altro se si ottengono stimatori di $\boldsymbol{\beta}$ che hanno la matrice di varianza covarianza “più piccola” o una matrice di informazione “più grande”.

Se $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$ e si hanno osservazioni indipendenti, le stime di massima verosimiglianza coincidono con le stime dei minimi quadrati. Inoltre la matrice di informazione del disegno $\xi = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, che abbiamo definito come $\mathbf{X}^t \mathbf{X}$, coincide con la matrice di informazione di Fisher. Nel caso in cui la distribuzione della variabile di risposta non è normale, supponendo le osservazioni indipendenti e sotto condizioni di regolarità, la matrice di varianza covarianza degli stimatori di massima verosimiglianza di $\boldsymbol{\beta}$ coincide asintoticamente con l'inversa della matrice di informazione di Fisher.

Le matrici di informazione possono essere confrontate mediante particolari criteri. Presentiamo ora alcuni tra i più diffusi criteri di ottimalità:

- A-ottimalità

se tutti i parametri sono stimabili, un possibile criterio è minimizzare la varianza

totale del vettore $\hat{\boldsymbol{\beta}}$ delle stime dei minimi quadrati. Nel caso di n osservazioni e matrice di disegno \mathbf{X} , il disegno ξ si dice A-ottimo se minimizza in Ξ

$$\sigma^2 \text{tr}[(\mathbf{X}^t \mathbf{X})^{-1}]. \quad (1.7)$$

Questo è anche detto *criterio della traccia*;

- c-ottimalità

supponiamo di essere interessati alla stima di una combinazione lineare dei parametri, di solito un particolare contrasto, che indichiamo con $\mathbf{c}^t \boldsymbol{\beta}$. Il disegno ξ è c-ottimo se minimizza in Ξ :

$$V(\mathbf{c}^t \hat{\boldsymbol{\beta}}) = \sigma^2 \mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c} = \sigma^2 \text{tr}[\mathbf{c}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c}] = \sigma^2 \text{tr}[(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{c} \mathbf{c}^t]; \quad (1.8)$$

- L-ottimalità

se ci interessa un vettore di s funzioni simultanee, sia esso $\mathbf{K}^t \boldsymbol{\beta}$ ove \mathbf{K} è una matrice $s \times p$, minimizziamo:

$$\sigma^2 \text{tr}[\mathbf{K}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{K}] = \sigma^2 \text{tr}[(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{K} \mathbf{K}^t] = \sigma^2 \text{tr}[(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{L}], \quad (1.9)$$

con \mathbf{L} matrice $p \times p$ semidefinita positiva opportuna. Questo criterio viene chiamato ottimalità lineare;

- D-ottimalità

il criterio più comunemente usato è senza dubbio quello della D-ottimalità. Un disegno ξ si dice D-ottimo se minimizza in Ξ

$$\det[\sigma^2 (\mathbf{X}^t \mathbf{X})^{-1}] = \sigma^{2p} \det[(\mathbf{X}^t \mathbf{X})^{-1}]. \quad (1.10)$$

Essendo $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$, il disegno D-ottimo massimizza il determinante della matrice di informazione del disegno oppure, il che è equivalente, massimizza il logaritmo del determinante di tale matrice

$$\lg[\det(\mathbf{X}^t \mathbf{X})]. \quad (1.11)$$

Sotto ipotesi di normalità, la D-ottimalità è equivalente a minimizzare il volume dell'ellissoide di confidenza per $\boldsymbol{\beta}$ a un dato livello. Questo criterio ha il vantaggio di essere invariante rispetto a riparametrizzazioni del modello, ad esempio cambiamenti di scala;

- G-ottimalità

se vogliamo predire $E(Y_x)$ su tutto lo spazio di disegno \mathcal{X} , ci interessa la varianza della risposta stimata in \mathbf{x} , ovvero $V(\hat{Y}_x) = \sigma^2 \mathbf{f}(\mathbf{x})^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{f}(\mathbf{x})$. Possiamo quindi minimizzare la varianza massima della previsione:

$$\min_{\xi} \max_{\mathbf{x} \in \mathcal{X}} \mathbf{f}(\mathbf{x})^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{f}(\mathbf{x}). \quad (1.12)$$

Questo criterio viene chiamato G-ottimalità ed è anch'esso indipendente dalla riparametrizzazione.

Fino adesso abbiamo definito un disegno come un insieme di punti $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ non necessariamente distinti, con $\mathbf{x}_i \in \mathcal{X}$, $i = 1, \dots, n$. Se invece consideriamo solo i punti di disegno effettivamente distinti possiamo descrivere un disegno ξ come un insieme di punti $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ replicati rispettivamente r_1, \dots, r_m volte, con $\sum_i r_i = n$. Quindi, in caso di osservazioni indipendenti, la matrice d'informazione può essere scritta come:

$$\sum_i \frac{r_i}{n} \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^t, \quad (1.13)$$

e nella (1.13) le quantità

$$\frac{r_i}{n}, \quad i = 1, \dots, m$$

possono essere descritte come porzioni del totale delle osservazioni da effettuare sui punti $\mathbf{x}_i, i = 1, \dots, m$. Pertanto, ad ogni disegno ξ può essere associata una misura discreta φ sullo spazio \mathcal{X} tale che $\varphi(\mathbf{x}_j) = 0$ se non devono essere effettuate osservazioni nel punto \mathbf{x}_j e $\varphi(\mathbf{x}_j) = r_i/n$ se devono essere eseguite r_i osservazioni nel punto \mathbf{x}_j . L'insieme dei punti con $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ con porzioni (dette anche *pesi*) rispettivamente r_i/n viene definito *supporto* del disegno. Per un disegno con n osservazioni, φ assume solo valori che sono multipli di $1/n$ e si definisce *disegno esatto*. Togliendo invece questo tipo di vincolo, si può estendere l'idea in questione ad una qualunque misura continua. Sia Φ l'insieme di tutte le misure di probabilità φ sulla σ -algebra di Borel su \mathcal{X} per cui

$$\varphi(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in \mathcal{X}, \quad \int_{\mathcal{X}} \varphi(d\mathbf{x}) = 1.$$

Ogni misura $\varphi \in \Phi$ definita come sopra rappresenta un disegno sullo spazio \mathcal{X} chiamato *disegno continuo*. La matrice di informazione per il disegno φ viene definita come:

$$\mathbf{M}(\varphi) = \int_{\mathcal{X}} \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^t \varphi(d\mathbf{x}).$$

In aggiunta, si può estendere la definizione di informazione di Fisher al caso di un disegno continuo ponendo

$$\mathbf{I}(\boldsymbol{\beta}; \varphi) = \int_{\mathcal{X}} \mathbf{I}(\boldsymbol{\beta}; \mathbf{x}) \varphi(d\mathbf{x}),$$

dove $\mathbf{I}(\boldsymbol{\beta}; \mathbf{x})$ è l'informazione per $\boldsymbol{\beta}$ in un punto di disegno \mathbf{x} .

Il disegno continuo non avrà in generale un'interpretazione diretta in termini di esperimenti reali ma deve essere visto come un artificio matematico che serve per sfruttare alcune proprietà fondamentali delle matrici di informazione. Ogni soluzione continua servirà solo come approssimazione di un vero piano di esperimento, cioè di un disegno esatto. I criteri di ottimalità precedentemente definiti possono essere estesi per un disegno continuo come segue:

- A-ottimalità: $\min_{\varphi} \text{tr}[\mathbf{M}(\varphi)^{-1}]$;
- c-ottimalità: $\min_{\varphi} \text{tr}[\mathbf{M}(\varphi)^{-1} \mathbf{c}\mathbf{c}^t]$;
- L-ottimalità: $\min_{\varphi} \text{tr}[\mathbf{M}(\varphi)^{-1} \mathbf{L}]$;
- D-ottimalità: $\min_{\varphi} \lg[\det(\mathbf{M}(\varphi))^{-1}] \equiv \max_{\varphi} \lg[\det(\mathbf{M}(\varphi))]$;
- G-ottimalità: $\min_{\varphi} \max_{\mathbf{x} \in \mathcal{X}} \mathbf{f}(\mathbf{x})^t (\mathbf{M}(\varphi))^{-1} \mathbf{f}(\mathbf{x})$.

Uno strumento fondamentale per affrontare l'ottimizzazione delle funzioni criterio è il cosiddetto Teorema di Equivalenza Generale di Kiefer e Wolfowitz (1960):

Teorema 1.4 *Le seguenti 3 condizioni sono equivalenti:*

1. φ^* massimizza $\det(\mathbf{M}(\varphi))$ nella classe di tutti i disegni continui su \mathcal{X} (D-ottimalità);
2. φ^* minimizza $\max_{\mathbf{x} \in \mathcal{X}} \mathbf{f}(\mathbf{x})^t (\mathbf{M}(\varphi))^{-1} \mathbf{f}(\mathbf{x})$ nella classe di tutti i disegni continui su \mathcal{X} (G-ottimalità);
3. $\max_{\mathbf{x} \in \mathcal{X}} \mathbf{f}(\mathbf{x})^t (\mathbf{M}(\varphi^*))^{-1} \mathbf{f}(\mathbf{x}) = p$, con p pari al numero dei parametri sconosciuti.

1.2.2 Problemi non Lineari

Anche se il modello lineare gode di grande popolarità, vi sono molti casi in cui una o più delle ipotesi di tale modello vengono meno, ad esempio:

- vogliamo stimare una funzione non lineare nei parametri;

- la risposta attesa è una funzione non lineare di parametri sconosciuti;
- la varianza delle osservazioni non è costante;
- le osservazioni sono correlate.

Un aspetto che accomuna i problemi di cui sopra è che la determinazione del disegno ottimo dipende quasi sempre dai parametri non noti. Vediamo alcuni esempi:

- Supponiamo di voler pianificare un esperimento per il modello lineare (1.1) ma di essere interessati alla stima di k funzioni non lineari dei parametri, ossia

$$g(\boldsymbol{\beta}) = (g_1(\boldsymbol{\beta}), \dots, g_k(\boldsymbol{\beta}))^t.$$

Supponiamo inoltre che le componenti di $g(\boldsymbol{\beta})$ siano funzioni di classe C^1 . Fissato un disegno φ , per osservazioni indipendenti si ha che le stime di massima verosimiglianza di $g(\boldsymbol{\beta})$ hanno matrice di varianza covarianza che asintoticamente coincide con l'inversa della matrice di informazione di Fisher che dipende dai parametri $\boldsymbol{\beta}$ e per la quale vale:

$$\mathbf{I}(\boldsymbol{\beta}; \varphi) = \nabla_g^t \left(\int \mathbf{I}(\boldsymbol{\beta}; \mathbf{x}) \varphi(d\mathbf{x}) \right) \nabla_g, \quad (1.14)$$

dove ∇_g è la matrice $p \times k$ con generico elemento i, j dato da $\partial g_j(\boldsymbol{\beta}) / \partial \beta_i$.

- Vogliamo adesso pianificare un esperimento ottimo per il seguente modello di regressione non lineare:

$$E(Y|\mathbf{x}) = \tau(\mathbf{x}, \boldsymbol{\beta}) \quad V(Y|\mathbf{x}) = \sigma^2, \quad (1.15)$$

in cui τ è una qualunque funzione di $\mathbf{x} \in \mathcal{X}$ e di $\boldsymbol{\beta}$ con Y distribuito normalmente. Se la funzione τ è derivabile parzialmente rispetto a β_i , $i = 1, \dots, p$, Box e Lucas (1959) suggerirono di rendere il problema del disegno equivalente a quello per il modello (1.4) approssimando la funzione τ con una funzione lineare sviluppandola in serie di Taylor al primo ordine intorno a una stima iniziale di $\boldsymbol{\beta}$. Il modello (1.15) quindi viene approssimato da un modello lineare nei parametri per il quale si può applicare la teoria vista precedentemente. Ovviamente però l'approssimazione dipende dalla stima iniziale di $\boldsymbol{\beta}$.

- Supponiamo ora di voler individuare un esperimento ottimo per un modello lineare generalizzato:

$$E(Y|\mathbf{x}) = h(\mathbf{f}(\mathbf{x})^t \boldsymbol{\beta}), \quad (1.16)$$

in cui $E(Y|\mathbf{x})$ può essere collegata al predittore lineare $\mathbf{f}(\mathbf{x})^t\boldsymbol{\beta}$ mediante una funzione link h monotona e differenziabile. In generale, il problema del disegno per questa classe di modelli può essere ricondotto a forma canonica (Ford et al, 1992). Mediante una trasformazione lineare e non singolare, lo spazio di disegno può essere trasformato in un nuovo spazio che dipende da $\boldsymbol{\beta}$. Lo scopo di questa trasformazione è quello di poter lavorare nelle nuove variabili ottenute, che sono opportune combinazioni lineari delle variabili di disegno originarie, in modo da non dover considerare nella fase di soluzione del problema la dipendenza esplicita della matrice di informazione del disegno dai parametri $\boldsymbol{\beta}$. Per i modelli lineari generalizzati, l'informazione di Fisher per $\boldsymbol{\beta}$ relativa al disegno φ ha la forma:

$$\mathbf{I}(\boldsymbol{\beta}; \varphi) = \int \omega(\mathbf{f}(\mathbf{x})^t\boldsymbol{\beta})\mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^t\varphi(d\mathbf{x}), \quad (1.17)$$

con $\omega(\cdot)$ funzione nota.

- Prendiamo adesso in considerazione il seguente modello di regressione

$$E(Y|\mathbf{x}) = \tau(\mathbf{x}, \boldsymbol{\beta}) \quad V(Y|\mathbf{x}) = \sigma^2(\mathbf{x}, \boldsymbol{\beta}), \quad (1.18)$$

in cui τ e σ^2 sono funzioni qualunque che dipendono da \mathbf{x} e dai parametri $\boldsymbol{\beta}$ che vogliamo stimare, avendo raccolto dei dati tramite un esperimento. Supponendo di conoscere la distribuzione di Y , che in genere non è normale, e assumendo le osservazioni indipendenti, sotto ipotesi di regolarità la matrice di varianza covarianza delle stime di massima verosimiglianza di $\boldsymbol{\beta}$ può, anche in questo caso, essere approssimata per grandi campioni dall'inversa della matrice di informazione di Fisher che però dipende dai parametri $\boldsymbol{\beta}$.

- Nel caso di pianificazione sperimentale per errori correlati, il problema riguarda sia il disegno che l'analisi dell'esperimento, naturalmente non quando la matrice di varianza covarianza è nota a meno di una costante, il che è equivalente al modello lineare tradizionale con le stime dei minimi quadrati ponderati, ma sotto l'ipotesi più realistica che tale matrice sia solo parzialmente nota, nel senso che si conosce il modo in cui dipende da certi parametri. Talvolta la correlazione è dovuta a prossimità spaziale delle unità sperimentali. Un caso particolare legato a questa problematica sarà trattato nel quarto capitolo di questa tesi.

In genere si chiama *non lineare* ogni problema di disegno in cui la matrice di informazione dipende dai parametri β . E' possibile, in teoria, risolvere anche in questi casi il problema del disegno e considerare i criteri di ottimalità precedentemente definiti. In genere però la soluzione dipenderà dai parametri stessi, dando luogo a disegni *localmente ottimi*. E' chiaro che non è questa la soluzione: vogliamo disegnare un esperimento per la stima dei parametri ma la soluzione dipende dai parametri stessi. In realtà esistono una serie di motivi per i quali può avere senso ricercare tali disegni, tra i quali:

- se si hanno informazioni sui parametri β , ad esempio da esperimenti precedenti, una soluzione è quella di basare l'esperimento sul piano che risulta ottimo relativamente a un valore dei parametri ipotizzato al meglio dalle conoscenze di partenza;
- disegni localmente ottimi possono essere utilizzati in piani sperimentali a più stadi, cioè il piano sperimentale all' i -esimo stadio può essere un disegno localmente ottimo in cui i valori di β sono approssimati dalle stime ottenute nei piani precedenti.

1.3 Piani Sequenziali

In ambito non lineare la metodologia sequenziale assume un'importanza fondamentale, in quanto rappresenta una soluzione naturale ai problemi di ottimalità locale. Infatti, attraverso l'utilizzo di procedure sequenziali è possibile ad ogni passo sfruttare le risposte osservate per stimare i parametri incogniti del modello, e quindi il disegno ottimo da essi dipendente. La pianificazione di un esperimento sequenziale viene scelta a priori dallo sperimentatore e comprende l'adozione del piano di campionamento delle unità statistiche, la scelta di una regola di determinazione delle assegnazioni e l'utilizzo di un'eventuale regola d'arresto. Per di più, spesso si è interessati ad ottimizzare l'esperimento rispetto a opportuni criteri di ottimalità, ad esempio la D-ottimalità, volti a formalizzare in modo matematico gli obiettivi sperimentali prefissati.

1.3.1 Procedura Sequenziale di Chaudhuri e Mykland

Nel campo degli esperimenti sequenziali per la stima di un modello in ambito non lineare, esiste una procedura suggerita da Chaudhuri e Mykland (1993) la quale può essere descritta come segue:

Sia n il numero totale dei punti di disegno e $\hat{\beta}_i$ lo stimatore dei parametri β basato

sulle osservazioni $(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_i, Y_i)$. Inoltre, indicheremo con \mathcal{R} lo spazio dove varia la variabile di risposta Y , con $I(\boldsymbol{\beta}; \mathbf{x})$ l'informazione per $\boldsymbol{\beta}$ in un punto di disegno \mathbf{x} e con $F_Y(y \mid \boldsymbol{\beta}, \mathbf{x})$ la funzione di probabilità (o densità nel caso di variabile di risposta continua) della distribuzione condizionale di Y dato \mathbf{x} . Infine, la misura del conteggio (o misura di Lebesgue nel caso in cui \mathcal{R} è un intervallo dell'asse dei reali) verrà indicata con λ .

FASE PRELIMINARE:

Vengono scelti $m < n$ punti di disegno $\mathbf{x}_1, \dots, \mathbf{x}_m$, nella maniera che il ricercatore ritiene più opportuna. Ad esempio, nel caso di assenza totale di informazione a priori per quanto riguarda i parametri nel modello, i punti possono essere distribuiti in modo equispaziato nel dominio se è un compatto. Successivamente, sulla base delle osservazioni ottenute $(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_m, Y_m)$ vengono stimati i parametri utilizzando un qualunque stimatore consistente, $\hat{\boldsymbol{\beta}}_m$.

FASE SEQUENZIALE:

1. *Esperimento*: la scelta dei punti di osservazione successivi viene effettuata attraverso un meccanismo basato sul criterio della D-ottimalità, ovvero si sceglie il punto di disegno che massimizza il determinante della matrice di informazione. Tale matrice dipende dai valori dei parametri e la si valuta nelle loro stime correnti;
2. *aggiornamento delle stime*: sulla base delle osservazioni ottenute precedentemente (incluse quelle fornite dalla fase preliminare) vengono stimati i parametri attraverso il metodo della massima verosimiglianza e si ritorna al passo 1 avendo aggiornato le stime.

Chaudhuri e Mykland (1993) hanno dimostrato che, sotto certe condizioni di regolarità e convergenza, gli stimatori di massima verosimiglianza ottenuti attraverso la suddetta metodologia sono consistenti ed asintoticamente normali. Tuttavia, tali condizioni sono numerose e spesso restrittive e difficili da verificare:

1. Lo spazio di disegno \mathcal{X} è un insieme compatto;
2. lo spazio parametrico \mathcal{B} è un sottoinsieme aperto e convesso di \mathbb{R}^d ;
3. l'informazione $\mathbf{I}(\boldsymbol{\beta}; \xi)$ può essere espressa nella forma $\mathbf{I}(\boldsymbol{\beta}; \xi) = \mathbf{V}(\boldsymbol{\beta}; \xi)\mathbf{V}(\boldsymbol{\beta}; \xi)^t$, con \mathbf{V} : funzione continua congiuntamente in entrambe le sue variabili $\boldsymbol{\beta}$ e \mathbf{x} ;

4. m tende all'infinito per n che tende all'infinito. Inoltre, i punti di disegno della fase preliminare sono scelti in modo che l'autovalore più piccolo della matrice $m^{-1} \sum_{i=1}^m \mathbf{I}(\boldsymbol{\beta}; \mathbf{x}_i)$ non si avvicina a zero per $m \rightarrow \infty$ per ogni $\boldsymbol{\beta} \in \mathcal{B}$;

5. la frazione m/n tende a zero per n che tende all'infinito;

6. per ogni $\epsilon > 0$, $\max_{m \leq i < n} P_{\boldsymbol{\beta}}(|\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}| > \epsilon) \rightarrow 0$ per $n \rightarrow \infty$;

7. per $m < k < n$, sia

$$U_k = \prod_{i=m+1}^k \det \left\{ \sum_{r=1}^i \mathbf{I}(\hat{\boldsymbol{\beta}}_{i-1}; \mathbf{x}_r) \right\} \cdot \det \left\{ \sum_{r=1}^i \mathbf{I}(\hat{\boldsymbol{\beta}}_i; \mathbf{x}_r) \right\}^{-1}.$$

Allora, per ogni $\epsilon > 0$, $\max_{m < k < n} P_{\boldsymbol{\beta}}(U_k > 1 + \epsilon) \rightarrow 0$ per $n \rightarrow \infty$;

8. il supporto di $F_Y(y | \boldsymbol{\beta}, \mathbf{x})$ non dipende né da $\boldsymbol{\beta}$ né da \mathbf{x} . Inoltre, per ogni $\mathbf{x} \in \mathcal{X}$ e $y \in \mathcal{R}$, $\lg\{F_Y(y | \boldsymbol{\beta}, \mathbf{x})\}$ è tre volte differenziabile con continuità in $\boldsymbol{\beta}$;

9. sia $\nabla \lg\{F_Y(y | \boldsymbol{\beta}, \mathbf{x})\} = \mathbf{G}(y, \boldsymbol{\beta}, \mathbf{x})$, il gradiente ottenuto calcolando le derivate parziali di primo ordine di $\lg\{F_Y(y | \boldsymbol{\beta}, \mathbf{x})\}$ rispetto a $\boldsymbol{\beta}$. Allora

$$\int_{\mathcal{R}} \mathbf{G}(y, \boldsymbol{\beta}, \mathbf{x}) F_Y(y | \boldsymbol{\beta}, \mathbf{x}) \lambda(dy) = 0, \sup_{\mathbf{x} \in \mathcal{X}} \int_{\mathcal{R}} |\mathbf{G}(y, \boldsymbol{\beta}, \mathbf{x})|^{2+t} F_Y(y | \boldsymbol{\beta}, \mathbf{x}) \lambda(dy) < \infty$$

per qualche $t > 0$;

10. sia $\mathbf{H}(y, \boldsymbol{\beta}, \mathbf{x})$ la matrice Hessiana di dimensione $d \times d$ ottenuta calcolando le derivate parziali di secondo ordine di $\lg\{F_Y(y | \boldsymbol{\beta}, \mathbf{x})\}$ rispetto a $\boldsymbol{\beta}$. Allora $\mathbf{H}(y, \boldsymbol{\beta}, \mathbf{x})$ soddisfa

$$\begin{aligned} \int_{\mathcal{R}} \mathbf{H}(y, \boldsymbol{\beta}, \mathbf{x}) F_Y(y | \boldsymbol{\beta}, \mathbf{x}) \lambda(dy) &= - \int_{\mathcal{R}} \{\mathbf{G}(y, \boldsymbol{\beta}, \mathbf{x})\} \{\mathbf{G}(y, \boldsymbol{\beta}, \mathbf{x})\}^t F_Y(y | \boldsymbol{\beta}, \mathbf{x}) \lambda(dy) \\ &= -\mathbf{I}(\boldsymbol{\beta}; \xi) \end{aligned}$$

e

$$\sup_{\mathbf{x} \in \mathcal{X}} \int_{\mathcal{R}} |\mathbf{H}(y, \boldsymbol{\beta}, \mathbf{x})|^2 F_Y(y | \boldsymbol{\beta}, \mathbf{x}) \lambda(dy) < \infty;$$

11. per ogni $\boldsymbol{\beta} \in \mathcal{B}$, esiste un intorno aperto $N(\boldsymbol{\beta})$ di $\boldsymbol{\beta}$ e una variabile aleatoria non negativa $\mathbf{K}(y, \boldsymbol{\beta}, \mathbf{x})$ tale che $\sup_{\mathbf{x} \in \mathcal{X}} \int_{\mathcal{R}} \mathbf{K}(y, \boldsymbol{\beta}, \mathbf{x}) F_Y(y | \boldsymbol{\beta}, \mathbf{x}) \lambda(dy) < \infty$ ed ognuna delle derivate parziali di $\lg\{F_Y(y | \boldsymbol{\beta}', \mathbf{x})\}$ di terzo ordine rispetto a $\boldsymbol{\beta}'$ è dominata da $\mathbf{K}(y, \boldsymbol{\beta}, \mathbf{x})$ per ogni $\boldsymbol{\beta}' \in N(\boldsymbol{\beta})$.

Come è chiaro, queste condizioni non sono di facile verificabilità. Quindi si tratta di trovare, caso per caso, esperimenti particolari la cui convergenza può essere verificata in maniera più diretta.

Capitolo 2

Modelli per Computer

Experiments

2.1 Esperimenti Simulati

La difficoltà, talvolta impossibilità, di effettuare sperimentazione fisica a causa dell'elevato dispendio di risorse temporali ed economiche risulta uno dei motivi fondamentali per cui si ricorre agli esperimenti fatti con l'uso del calcolatore, chiamati, appunto, computer experiments (CE): all'osservazione di un fenomeno nella realtà viene sostituito l'output di un simulatore. In aggiunta, negli anni passati, numerosi modelli teorici sono stati esaminati soltanto in alcune circostanze speciali. Ad esempio, nella fisica della materia condensata, un modello per le forze intermolecolari in un materiale specifico, veniva studiato in un cristallo perfetto dove tutti gli atomi occupano le corrette posizioni reticolari nella struttura cristallina. Anche se il cristallo perfetto può esistere solo ipoteticamente, l'assunzione di tale approssimazione era necessaria per la trattazione del problema e per i vari calcoli che dovevano essere effettuati per lo studio del fenomeno. L'avvento delle nuove tecnologie e l'utilizzo dei computer experiments hanno modificato il precedente scenario rendendo possibile l'abbandono di ipotesi semplificate e troppo spesso irrealistiche, introducendo complessità al modello e aprendo in questo modo la strada verso una comprensione migliore degli esperimenti reali. I calcolatori permettono, quindi, di simulare complessi fenomeni fisici tramite i cosiddetti codici o computer models, cioè modelli matematici deterministici i quali vengono implementati al computer e permettono di esplorare la complessa relazione tra fattori di input e variabili di output.

Tuttavia, dovendo descrivere in modo accurato un sistema fisico, un modello di simulazione tende ad essere esso stesso complesso. In molti casi, infatti, il modello può essere riguardato come la soluzione di un insieme di numerose equazioni lineari, non lineari, ordinarie e/o differenziali ed è, in genere, non esprimibile sotto una forma analitica trattabile. In aggiunta, il numero di fattori di input da considerare può risultare elevato per cui in molte situazioni risulta impossibile esplorare l'intero dominio. La suddetta complessità rende le simulazioni costose sia da un punto di vista economico che temporale. Per questi motivi si usa ricorrere ai metamodelli, cioè a modelli molto più semplici che costituiscono una valida approssimazione del modello originale.

2.2 Metamodelli

Il concetto di metamodello è stato formulato per la prima volta da Blanning (1975), il quale propose di costruire dei *modelli di modelli di simulazione*. Nella letteratura, il metamodello è anche noto come *modello approssimato*, *surrogato* o *emulatore*. Si tratta di un modello creato al fine di ridurre la complessità del modello di partenza, il quale viene generalmente chiamato *codice*, *modello originale*, *simulatore* o *modello di simulazione*. L'obiettivo del surrogato è quindi riprodurre il modello originale, il quale, a sua volta, simula il comportamento di un sistema reale. Esistono varie tecniche per la costruzione dei metamodelli. Una di esse, ampiamente condivisa nella letteratura, è la metodologia Kriging la quale sarà l'oggetto principale di questo capitolo.

2.2.1 Obiettivo del Metamodello

Un punto cruciale per la costruzione degli emulatori è quello della specificazione dello scopo che si vuole raggiungere tramite il loro utilizzo. Da questa scelta dipendono molteplici caratteristiche assunte dal metamodello. Di seguito elenchiamo alcune tra le diverse finalità:

- **previsione:** il metamodello viene costruito per effettuare delle previsioni del sistema fisico;
- **ottimizzazione:** il surrogato viene utilizzato al posto del simulatore per risolvere un problema di ottimizzazione riguardante il sistema reale.

Le categorie di cui sopra non sono necessariamente esclusive tra loro: anche se il meta-modello viene identificato con uno specifico obiettivo, esso può essere comunque usato anche per altri scopi; ad esempio, un emulatore costruito per un problema di ottimizzazione può fornire utili indicazioni per la comprensione globale del modello originale e, di conseguenza, del sistema reale.

2.2.2 Procedura di Sviluppo del Metamodello

La costruzione di un emulatore può essere descritta in modo schematico attraverso i seguenti passaggi (per approfondimenti sulla procedura per la costruzione dei metamodelli si veda Kleijnen e Sargent, 2000):

- 1) *Determinazione dello scopo del metamodello*: prevede di specificare l'obiettivo principale per cui si ricorre all'uso del surrogato;
- 2) *Identificazione della variabile di output*: i problemi per cui si ricorre all'uso del simulatore sono assai complessi e pertanto, nella maggior parte dei casi, il simulatore stesso è caratterizzato da numerose variabili di output. Tuttavia, è prassi sviluppare metamodelli diversi ciascuno con una sola variabile di risposta, anziché un solo surrogato complesso. Quindi, fissato lo scopo del metamodello, è possibile determinare la variabile di risposta e la sua unità di misura;
- 3) *Identificazione delle variabili di input (screening)*: lo screening di un metamodello riguarda la determinazione dei fattori più importanti. Di conseguenza, devono essere selezionate le variabili di ingresso giudicate più rilevanti, definendo l'insieme dei valori possibili che esse possono assumere (dominio o spazio di disegno);
- 4) *Specificazione dell'accuratezza richiesta*: questa caratteristica è strettamente legata alla finalità del surrogato (passo 1). Ad esempio, se l'obiettivo principale consiste nell'ottimizzazione di un sistema, non si può richiedere elevata complessità, dal momento che gli algoritmi di ottimizzazione necessitano di modelli semplici;
- 5) *Scelta del tipo del surrogato*: in questo passo si deve selezionare la tipologia del metamodello tra i tanti proposti in letteratura (modelli polinomiali e razionali, splines, Fourier, reti neurali, Kriging...);
- 6) *Definizione del piano sperimentale*: è necessario specificare l'insieme dei punti (chiamato disegno) ciascuno dei quali costituito da una certa combinazione dei diversi

fattori sotto controllo sperimentale, in corrispondenza dei quali verranno effettuate le simulazioni dal modello originale, cioè dal simulatore, per stimare i parametri dei metamodelli parametrici. Due sono gli approcci solitamente utilizzati per la scelta del disegno sperimentale. Il primo prevede l'utilizzo dei cosiddetti disegni Space-Filling, ossia procedure che sono spesso basate su metodi di campionamento e mirano a distribuire i punti del dominio in modo equispaziato motivate dall'esigenza di osservare la risposta lungo l'intero spazio di disegno; il secondo si basa sulla teoria dei disegni ottimi;

- 7) *Identificazione del metamodello*: sulla base delle osservazioni simulate in corrispondenza dei punti di disegno ottenuti nel passo precedente, viene identificato il modello approssimato, ossia si stimano i parametri d'interesse dell'emulatore scelto al passo 5;
- 8) *Validazione del modello approssimato*: in ultima istanza si deve controllare l'accuratezza del metamodello così costruito attraverso dati prodotti dal modello originale.

2.3 Modelli basati sulle Funzioni-base

Siccome l'output dei modelli di simulazione è deterministico, la costruzione di un surrogato $g(\cdot)$ può essere vista come un problema di interpolazione. La maggior parte dei metamodelli che seguono questa filosofia hanno la seguente forma:

$$g(\mathbf{x}) = \sum_{j=0}^L B_j(\mathbf{x})\beta_j \quad (2.1)$$

dove i β_j sono dei parametri da stimare e $\{B_0(\mathbf{x}), \dots, B_L(\mathbf{x})\}$ è l'insieme delle funzioni-base che possono essere di tipo polinomiale, splines, Fourier, razionali....Per la descrizione dei metamodelli appartenenti alla classe (2.1) utilizzeremo la seguente notazione matriciale assumendo di effettuare n osservazioni

$$\mathbf{B} = \begin{pmatrix} B_0(\mathbf{x}_1) & \dots & B_L(\mathbf{x}_1) \\ B_0(\mathbf{x}_2) & \dots & B_L(\mathbf{x}_2) \\ \dots & \dots & \dots \\ B_0(\mathbf{x}_n) & \dots & B_L(\mathbf{x}_n) \end{pmatrix} \quad (2.2)$$

$$\mathbf{y} = (y(\mathbf{x}_1), \dots, y(\mathbf{x}_n))^t \quad \boldsymbol{\beta} = (\beta_0, \dots, \beta_L)^t \quad \mathbf{b}(\mathbf{x}) = (B_0(\mathbf{x}), \dots, B_L(\mathbf{x}))^t$$

Per interpolare gli output osservati $y(\mathbf{x}_1), \dots, y(\mathbf{x}_n)$ corrispondenti a $\mathbf{x}_1, \dots, \mathbf{x}_n$ utilizzando le basi $B_0(\mathbf{x}), B_1(\mathbf{x}), \dots$ si deve prendere un L abbastanza grande tale che l'equazione

$$\mathbf{y} = \mathbf{B}\boldsymbol{\beta} \quad (2.3)$$

abbia soluzione. Potrebbe accadere che (2.3) risulti un sistema indeterminato. Per risolvere tale problematica è stato proposto da Rao (1973) e Antoniadis e Fan (2001) il cosiddetto *normalized method of frame* che consiste nello scegliere un insieme di parametri $\boldsymbol{\beta}$ che minimizzano $\|\boldsymbol{\beta}\|^2$. Li (2002) ha trovato un legame tra il metodo normalizzato di cui sopra e i minimi quadrati penalizzati. Egli, al posto della minimizzazione vincolata ha considerato la seguente espressione di minimizzazione:

$$\|\boldsymbol{\beta}\|^2 + \lambda \|\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\|^2, \quad (2.4)$$

dove $\|\cdot\|$ è la norma euclidea e λ un moltiplicatore di Lagrange. La (2.4) quindi può essere vista come una somma dei minimi quadrati penalizzati, ovvero

$$\frac{1}{2} \|\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\|^2 + \lambda \sum_{j=1}^L \beta_j^2, \quad (2.5)$$

dove λ è un parametro di regolarizzazione e il secondo addendo dell'espressione (2.5) è una particolare funzione di penalizzazione. La soluzione di (2.5) può essere espressa come

$$\widehat{\boldsymbol{\beta}}_\lambda = (\mathbf{B}^t \mathbf{B} + 2\lambda \mathbf{I}_L)^{-1} \mathbf{B}^t \mathbf{y}, \quad (2.6)$$

dove \mathbf{I}_L è la matrice identità L -dimensionale, $\widehat{\boldsymbol{\beta}}_\lambda$ si chiama stimatore di tipo *ridge* e λ parametro *ridge*. In questo modo si ottiene il metamodello $g_\lambda(\mathbf{x}) = \mathbf{b}(\mathbf{x})^t \widehat{\boldsymbol{\beta}}_\lambda$. La scelta di λ viene effettuata con l'obiettivo di garantire un opportuno bilanciamento tra la varianza e la distorsione dello stimatore ($\lambda = 0$ corrisponde alle stime dei minimi quadrati ordinari). Si possono avere delle funzioni di penalizzazione diverse da quella che appare in (2.5). Considerando i minimi quadrati penalizzati

$$Q(\boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\|^2 + n \sum_{j=0}^L p_\lambda(|\beta_j|), \quad (2.7)$$

dove $p_\lambda(\cdot)$ è una funzione di penalizzazione non-negativa e λ un parametro di regolarizzazione che potrebbe dipendere da n , alcune delle $p_\lambda(\cdot)$ proposte in letteratura sono:

$$(L_0) \quad p_\lambda(|\beta|) = \frac{1}{2} \lambda^2 I(|\beta| \neq 0), \text{ dove } I \text{ è la funzione indicatrice;}$$

$$(L_1) \quad p_\lambda(|\beta|) = \lambda |\beta|;$$

$$(L_q) \quad p_\lambda(|\beta|) = \lambda |\beta|^q, \quad q > 0.$$

2.3.1 Modelli Polinomiali

Come abbiamo già accennato, ci sono vari tipi di funzioni-base. Nei modelli polinomiali esse sono:

$$\begin{aligned} B_0(\mathbf{x}) &= 1, \quad B_1(\mathbf{x}) = x_1, \dots, \quad B_s(\mathbf{x}) = x_s, \\ B_{s+1}(\mathbf{x}) &= x_1^2, \dots, \quad B_{2s}(\mathbf{x}) = x_s^2, \\ B_{2s+1}(\mathbf{x}) &= x_1x_2, \dots, \quad B_{\frac{s(s+3)}{2}}(\mathbf{x}) = x_{s-1}x_s, \quad \dots \end{aligned} \quad (2.8)$$

Si noti che il numero delle funzioni-base di cui sopra cresce notevolmente con il numero delle variabili di input e il grado del polinomio. Perciò, è prassi l'uso di polinomi che non hanno un grado elevato. Ad esempio, i polinomi di secondo grado sono tra i più popolari nell'ambito dei computer experiments e il metodo dei minimi quadrati viene frequentemente usato per stimare β .

Le funzioni-base (2.8) non formano una base ortogonale per cui si può riscontrare il problema della collinearità, ossia correlazione tra i regressori. Per attenuare questo effetto negativo sono state introdotte delle funzioni-base polinomiali ortogonali per modellare i computer experiments (si veda An e Owen, 2001). Nella letteratura si possono trovare numerosi casi studio nei quali vengono sfruttati i metamodelli di questa classe (si veda ad esempio Bates et al, 2003).

2.3.2 Splines

E' prassi utilizzare i metamodelli polinomiali se si vuole avere un'idea del comportamento globale del modello vero. Tuttavia, quando il dominio dei fattori di input è vasto e il modello di simulazione molto complesso con numerosi punti di minimo e massimo, si possono sfruttare, al posto dei polinomi di grado elevato, le cosiddette funzioni spline. In generale, una spline è una funzione costituita da un insieme di polinomi raccordati tra loro, il cui scopo è interpolare in un intervallo un insieme di punti, detti nodi della spline, in modo da essere continua, almeno fino ad un dato ordine di derivate, in ogni punto dell'intervallo. In questo contesto, si usa sempre il metamodello dato dall'equazione (2.1) e si ricorre alle splines per la costruzione delle funzioni-base:

Definizione 2.1 Sia data $\{a = \kappa_1 < \kappa_2 < \dots < \kappa_K = b\}$ una suddivisione dell'intervallo $[a, b]$. Una *spline di grado p* con nodi nei punti κ_i con $i = 1, \dots, K$ è una funzione su $[a, b]$ indicata con $s_p(x)$ tale che, nell'intervallo $[a, b]$ si abbia:

- (i) In ogni sottointervallo $[\kappa_i, \kappa_{i+1}]$ con $i = 1, \dots, K - 1$ la funzione $s_p(x)$ è un polinomio di grado p ;
- (ii) la funzione $s_p(x)$ e le sue prime p derivate sono continue.

Nell'ambito dei computer experiments, se si considera una variabile di input unidimensionale, le funzioni-base che si ottengono utilizzando le splines hanno la seguente forma:

$$B_0(x) = 1, B_1(x) = x, B_2(x) = x^2, \dots, B_p(x) = x^p,$$

$$B_{p+1}(x) = (x - \kappa_1)_+^p, \dots, B_{p+K}(x) = (x - \kappa_K)_+^p \quad (2.9)$$

dove $\kappa_1, \dots, \kappa_K$ è l'insieme dei nodi nei quali la $(p + 1)$ -esima derivata di g è discontinua, e α_+ è la parte positiva di α . Pertanto, il metamodello ottenuto è una spline di ordine p :

$$s(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_p x^p + \sum_{k=1}^K \beta_{p+k} (x - \kappa_k)_+^p$$

Le funzioni-base multivariate si possono costruire a partire da quelle univariate tramite il prodotto tensore. Se si considera una variabile di input $\mathbf{x} = (x_1, \dots, x_s)^t$, la funzione-base s -dimensionale costruita tramite il prodotto tensore è della seguente forma

$$B_{r_1, \dots, r_s}(\mathbf{x}) = \prod_{j=1}^s B_{r_j}(x_j)$$

Il prodotto tensore costituisce uno strumento diretto per estendere le funzioni-base unidimensionali al caso multidimensionale. Tuttavia, la sua implementazione presenta molte difficoltà dal momento che il numero dei coefficienti da stimare cresce rapidamente al crescere della dimensione. In alternativa all'utilizzo del prodotto tensore, Friedman (1991) propose l'algoritmo MARS (multivariate adaptive regression splines). Per approfondimenti sull'algoritmo si veda il lavoro di cui sopra.

2.3.3 Fourier

Le funzioni-base polinomiali e splines sono quelle maggiormente usate in letteratura. Tuttavia, a partire dall'equazione (2.1) e sfruttando le funzioni-base di Fourier si possono approssimare delle funzioni periodiche. E' ben noto che

$$1, \cos(2\pi x), \sin(2\pi x), \dots, \cos(2k\pi x), \sin(2k\pi x), \dots$$

formano una base ortogonale in $[0, 1]$. Tramite il prodotto tensore si costruisce la base s -dimensionale e il metamodella ottenuto

$$\beta_0 + \sum_{i=1}^s \sum_{j=1}^m \{\alpha_{ij} \cos(2j\pi x_i) + \beta_{ij} \sin(2j\pi x_i)\}$$

viene chiamato *modello di regressione di Fourier*. Si veda Bates et al (1996) per l'uso delle funzioni-base di Fourier per modellare i computer experiments.

2.4 Random Field Regression Models

Finora ci siamo occupati della costruzione dei metamodelli visti come un problema di interpolazione. Tuttavia, tale tipologia di modellistica soffre di un problema di dimensionalità. Ad esempio, si consideri un surrogato basato su funzioni polinomiali; anche arrestandoci al secondo ordine e prendendo otto fattori di input, il metamodella assume la forma:

$$\beta_0 + \beta_1 x_1 + \dots + \beta_8 x_8 + \beta_9 x_1^2 + \dots + \beta_{16} x_8^2 + \beta_{17} x_1 x_2 + \dots + \beta_{44} x_7 x_8.$$

Pertanto i parametri da stimare sono ben 45! In generale, i metamodelli basati sulle funzioni-base, qualsiasi esse siano, sono competitivi solo in presenza di un numero limitato di variabili di ingresso in quanto in caso di elevata dimensionalità il numero dei coefficienti da stimare cresce molto rapidamente.

Una soluzione alternativa ampiamente condivisa in letteratura consiste nel modellizzare l'incertezza nella relazione tra fattori di input e risposta considerando l'output del modello come realizzazione di un processo stocastico. Una delle prime proposte e più significative all'utilizzo di questa metodologia nell'ambito della sperimentazione simulata è quella di Sacks et al (1989) dove viene assunto che ciascuna osservazione $y(\mathbf{x}) = f(x_1, \dots, x_s)$ effettuata nel punto \mathbf{x} , sia realizzazione di un processo stocastico $Y(\mathbf{x})$ modellato come

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^s \quad (2.10)$$

dove $\mu(\cdot)$ è il cosiddetto trend globale il quale viene solitamente modellato da $\mu(\mathbf{x}) = \sum_{j=1}^p f_j(\mathbf{x})\beta_j = \mathbf{f}(\mathbf{x})^t \boldsymbol{\beta}$, con $f_1(\cdot), \dots, f_p(\cdot)$ funzioni note, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ il vettore dei parametri incogniti e $Z(\mathbf{x})$ la componente stocastica che rappresenta lo scostamento della risposta $Y(\mathbf{x})$ dal trend. $Z(\mathbf{x})$ è un processo Gaussiano stazionario di media zero, varianza costante σ_Z^2 e funzione di correlazione (non negativa) tra due input \mathbf{x} e \mathbf{x}' data da

$$\mathcal{R}(Z(\mathbf{x}), Z(\mathbf{x}')) = \mathcal{R}(\|\mathbf{x} - \mathbf{x}'\|)$$

tale che

$$\mathcal{R}(Z(\mathbf{x}), Z(\mathbf{x}')) \rightarrow 1, \quad \text{per } \|\mathbf{x} - \mathbf{x}'\| \rightarrow 0 \quad \text{e} \quad (2.11)$$

$$\mathcal{R}(Z(\mathbf{x}), Z(\mathbf{x}')) \rightarrow 0, \quad \text{per } \|\mathbf{x} - \mathbf{x}'\| \rightarrow \infty \quad (2.12)$$

Osservazione 2.2 *La natura deterministica del problema, ovvero l'assenza della componente casuale dell'errore, viene sottolineata da (2.11).*

2.4.1 Modelli di Correlazione

La scelta della funzione di correlazione è essenziale in questo contesto. Di seguito elenchiamo alcuni modelli di correlazione sottolineando che viene assunto che essi soddisfino la cosiddetta “regola del prodotto delle correlazioni unidimensionali”, ossia

$$\mathcal{R}(\mathbf{x}, \mathbf{x}') = \prod_j R_j(x_j - x'_j)$$

Modello	\mathcal{R}_j
Esponenziale	$\exp[-\theta_j x_j - x'_j ^{\alpha_j}], \quad 0 < \alpha_j \leq 2$
Lineare	$\theta_j x_j - x'_j $
Gaussiano	$\exp[-\theta_j (x_j - x'_j)^2]$
Sferico	$1 - 1.5\xi_j + 0.5\xi_j^3, \quad \xi_j = \min\{1, \theta_j x_j - x'_j \}$
Cubico	$1 - 3\xi_j^2 + 2\xi_j^3, \quad \xi_j = \min\{1, \theta_j x_j - x'_j \}$

2.4.2 Previsori

Siano

$$Y(\mathbf{x}) = \sum_{j=1}^p f_j(\mathbf{x})\beta_j + Z(\mathbf{x}) = \mathbf{f}(\mathbf{x})^t \boldsymbol{\beta} + Z(\mathbf{x})$$

il Random Field Regression Model, $\mathbf{x}_1, \dots, \mathbf{x}_n$ gli input scelti,

$$\mathbf{Y}^n = (Y_1, \dots, Y_n)^t = (y(\mathbf{x}_1), \dots, y(\mathbf{x}_n))^t$$

gli output osservati e $y(\mathbf{x}_0) = Y_0$ l'output da predire. Infine, si denoti con \hat{Y}_0 un generico previsore di Y_0 . Con le ipotesi fatte precedentemente, il modello in questione può essere

descritto nel seguente modo

$$\begin{pmatrix} Y_0 \\ \mathbf{Y}^n \end{pmatrix} \sim \left[\begin{pmatrix} \mathbf{f}_0^t \\ \mathbf{F} \end{pmatrix} \beta, \sigma_Z^2 \begin{pmatrix} 1 & \mathbf{r}_0^t \\ \mathbf{r}_0 & \mathbf{R} \end{pmatrix} \right] \quad (2.13)$$

dove $\mathbf{f}_0 = \mathbf{f}(\mathbf{x}_0)$ è il vettore delle funzioni di regressione per Y_0 di dimensione $(p \times 1)$,

$$\mathbf{F} = (f_j(\mathbf{x}_i)) = \begin{pmatrix} \mathbf{f}_1(\mathbf{x}_1) & \dots & \mathbf{f}_p(\mathbf{x}_p) \\ \vdots & & \vdots \\ \mathbf{f}_1(\mathbf{x}_n) & \dots & \mathbf{f}_p(\mathbf{x}_n) \end{pmatrix}$$

è la matrice delle funzioni di regressione per i dati osservati di dimensione $(n \times p)$ e $\mathbf{r}_0 = (\mathcal{R}(\mathbf{x}_0 - \mathbf{x}_1), \dots, \mathcal{R}(\mathbf{x}_0 - \mathbf{x}_n))^t$ rappresenta il vettore $(n \times 1)$ -dimensionale delle correlazioni tra \mathbf{Y}^n e Y_0 . Infine \mathbf{R} , avente come generico elemento

$$\mathbf{R}_{i,j} = (\mathcal{R}(\mathbf{x}_i, \mathbf{x}_j)),$$

è la matrice delle correlazioni di \mathbf{Y}^n , e β e σ_Z^2 sono parametri non noti. Per quanto riguarda i previsori di Y_0 si devono distinguere due casi:

- (i) Funzione di correlazione nota;
- (ii) funzione di correlazione non nota.

Funzione di correlazione nota

Si dimostra che, nel caso in cui $\mathcal{R}(\cdot)$ è specificata, il BLUP (Best Linear Unbiased Predictor) per Y_0 è dato da

$$\hat{Y}_0 \equiv \mathbf{f}_0^t \hat{\beta} + \mathbf{r}_0^t \mathbf{R}^{-1} (\mathbf{Y}^n - \mathbf{F} \hat{\beta}) \quad (2.14)$$

dove $\hat{\beta} = (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathbf{R}^{-1} \mathbf{Y}^n$ è lo stimatore dei minimi quadrati generalizzati. Prima di dimostrare che (2.14) è il Best Linear Unbiased Predictor per questo modello, elenchiamo alcune delle sue proprietà:

- (i) \hat{Y}_0 interpola i dati $(\mathbf{x}_i, y(\mathbf{x}_i))$ per $1 \leq i \leq n$;
- (ii) \hat{Y}_0 è LUP (Linear Unbiased Predictor) per $Y(\mathbf{x}_0)$. Infatti, esso risulta una combinazione lineare delle osservazioni:

$$\hat{Y}_0 = \{\mathbf{f}_0^t (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathbf{R}^{-1} + \mathbf{r}_0^t \mathbf{R}^{-1} [\mathbf{I}_n - \mathbf{F} (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathbf{R}^{-1}]\} \mathbf{Y}^n \equiv \mathbf{a}_*^t \mathbf{Y}^n \quad (2.15)$$

dove \mathbf{a}_* viene definito da (2.15). Inoltre, \hat{Y}_0 è uno stimatore corretto:

$$E(\hat{Y}_0) = \mathbf{a}_*^t E(\mathbf{Y}^n) = \mathbf{a}_*^t \mathbf{F} \boldsymbol{\beta} = [\mathbf{f}_0^t \mathbf{I}_n + \mathbf{r}_0^t \mathbf{R}^{-1} (\mathbf{F} - \mathbf{F} \mathbf{I}_n)] \boldsymbol{\beta} = \mathbf{f}_0^t \boldsymbol{\beta} \quad (2.16)$$

Si dimostra ora che (2.14) è il Best Linear Unbiased Predictor rispetto alla famiglia (2.13). Abbiamo già fatto vedere che $\hat{Y}_0 = \mathbf{a}_*^t \mathbf{Y}^n$, ossia il fatto che esso risulta essere una combinazione lineare delle osservazioni. Inoltre, dalla proprietà della correttezza (2.16) si ottiene $\mathbf{F}^t \mathbf{a}_* = \mathbf{f}_0$. Si denotino adesso con $\mathbf{a}^t \mathbf{Y}^n$ l'insieme di tutti i previsori lineari e corretti di Y_0 e con $\mathbf{Z}^n = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))^t$ e $Z_0 = Z(\mathbf{x}_0)$ lo scostamento dal modello lineare di \mathbf{Y}^n e di Y_0 , rispettivamente. Fissati $\boldsymbol{\beta}$ e σ_Z^2 si ha

$$\begin{aligned} E\{(\mathbf{a}^t \mathbf{Y}^n - Y_0)^2\} &= E\{[\mathbf{a}^t (\mathbf{F} \boldsymbol{\beta} + \mathbf{Z}^n) - (\mathbf{f}_0^t \boldsymbol{\beta} + Z_0)]^2\} = E\{[(\mathbf{a}^t \mathbf{F} - \mathbf{f}_0^t) \boldsymbol{\beta} + \mathbf{a}^t \mathbf{Z}^n - Z_0]^2\} = \\ &= E\{\mathbf{a}^t \mathbf{Z}^n (\mathbf{Z}^n)^t \mathbf{a} - 2\mathbf{a}^t \mathbf{Z}^n Z_0 + Z_0^2\} = \sigma_Z^2 \mathbf{a}^t \mathbf{R} \mathbf{a} - 2\sigma_Z^2 \mathbf{a}^t \mathbf{r}_0 + \sigma_Z^2 = \sigma_Z^2 (\mathbf{a}^t \mathbf{R} \mathbf{a} - 2\mathbf{a}^t \mathbf{r}_0 + 1), \end{aligned}$$

dove la terza uguaglianza segue dal vincolo imposto dalla correttezza, ovvero $\mathbf{F}^t \mathbf{a} = \mathbf{f}_0$. Pertanto, il BLUP è dato da $\mathbf{a}_{**}^t \mathbf{Y}^n$ con \mathbf{a}_{**} che minimizza

$$\mathbf{a}^t \mathbf{R} \mathbf{a} - 2\mathbf{a}^t \mathbf{r}_0 \quad (2.17)$$

sotto il vincolo

$$\mathbf{F}^t \mathbf{a} = \mathbf{f}_0. \quad (2.18)$$

Il metodo dei moltiplicatori di Lagrange può essere applicato per minimizzare (2.17) soggetto alla condizione (2.18), ossia trovare il vettore $(\mathbf{a}, \boldsymbol{\lambda}) \in \mathbb{R}^{n+p}$ che minimizza

$$\mathbf{a}^t \mathbf{R} \mathbf{a} - 2\mathbf{a}^t \mathbf{r}_0 + 2\boldsymbol{\lambda}^t (\mathbf{F}^t \mathbf{a} - \mathbf{f}_0) \quad (2.19)$$

Calcolando il gradiente di (2.19) rispetto a $(\mathbf{a}, \boldsymbol{\lambda})$ e ponendolo uguale al vettore zero, si ottiene il seguente sistema

$$\mathbf{F}^t \mathbf{a} - \mathbf{f}_0 = \mathbf{0}$$

$$\mathbf{R} \mathbf{a} - \mathbf{r}_0 + \mathbf{F} \boldsymbol{\lambda} = \mathbf{0}$$

o, equivalentemente, in notazione matriciale

$$\begin{pmatrix} 0 & \mathbf{F}^t \\ \mathbf{F} & \mathbf{R} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{a} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_0 \\ \mathbf{r}_0 \end{pmatrix}$$

da cui si ottiene

$$\begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{a} \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{F}^t \\ \mathbf{F} & \mathbf{R} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{f}_0 \\ \mathbf{r}_0 \end{pmatrix} = \begin{pmatrix} -\mathbf{Q} & \mathbf{Q} \mathbf{F}^t \mathbf{R}^{-1} \\ \mathbf{R}^{-1} \mathbf{F} \mathbf{Q} & \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{F} \mathbf{Q} \mathbf{F}^t \mathbf{R}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{f}_0 \\ \mathbf{r}_0 \end{pmatrix} \quad (2.20)$$

dove $\mathbf{Q} = (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F})^{-1}$. Dopo aver fatto un paio di semplici passaggi algebrici, si ottiene il risultato desiderato.

Funzione di correlazione non nota

In questo contesto, l'idea principale consiste nel trovare un previsore di Y_0 tramite

$$\hat{Y}_0 = \mathbf{f}_0^t \hat{\boldsymbol{\beta}} + \hat{\mathbf{r}}_0^t \hat{\mathbf{R}}^{-1} (\mathbf{Y}^n - \mathbf{F} \hat{\boldsymbol{\beta}}) \quad (2.21)$$

dove $\hat{\boldsymbol{\beta}} = (\mathbf{F}^t \hat{\mathbf{R}}^{-1} \mathbf{F})^{-1} \mathbf{F}^t \hat{\mathbf{R}}^{-1} \mathbf{Y}^n$ e gli stimatori $\hat{\mathbf{R}}$ e $\hat{\mathbf{r}}_0$ vengono determinati per mezzo di uno stimatore della funzione di correlazione $\mathcal{R}(\cdot)$. Tali previsori vengono chiamati Empirical Best Linear Unbiased Predictor (EBLUP) di Y_0 anche se, nella maggior parte dei casi, essi non sono né corretti né combinazioni lineari delle osservazioni. Difatti, $\hat{\mathbf{R}} = \hat{\mathbf{R}}(\mathbf{Y}^n)$ e $\hat{\mathbf{r}}_0 = \hat{\mathbf{r}}_0(\mathbf{Y}^n)$ sono in genere non lineari in \mathbf{Y}^n .

Nella letteratura, è prassi assumere per gli stimatori della funzione di correlazione

$$\mathcal{R}(\cdot) = \mathcal{R}(\cdot | \boldsymbol{\psi})$$

con $\boldsymbol{\psi}$ un vettore di parametri. Ad esempio, la funzione di correlazione di tipo esponenziale

$$\mathcal{R}(\cdot | \boldsymbol{\psi}) = \prod_{j=1}^d \exp(-\theta_j |x_j - x'_j|^{\alpha_j})$$

ha d parametri di scala $\theta_1, \dots, \theta_d$ e d parametri smoothness $\alpha_1, \dots, \alpha_d$. Pertanto, $\boldsymbol{\psi}$ è un vettore di $2 \times d$ componenti. In questo caso, sia la matrice di correlazione $\mathbf{R} = \mathbf{R}(\boldsymbol{\psi})$ che il vettore \mathbf{r}_0 dipendono da $\boldsymbol{\psi}$. Di seguito descriviamo due metodi diversi per stimare il vettore $\boldsymbol{\psi}$ che portano a diversi stimatori EBLUP. Entrambe le tecniche assumono per la distribuzione condizionale

$$[\mathbf{Y}^n | \boldsymbol{\beta}, \sigma_Z^2, \boldsymbol{\psi}] \sim N_n[\mathbf{F} \boldsymbol{\beta}, \sigma_Z^2 \mathbf{R}]. \quad (2.22)$$

Massima Verosimiglianza

Lo stimatore di massima verosimiglianza (Maximum Likelihood Estimate) è la scelta più popolare per stimare $\boldsymbol{\psi}$. Sotto l'assunzione di normalità (2.22), la log verosimiglianza risulta (a meno di una costante)

$$l(\boldsymbol{\beta}, \sigma_Z^2, \boldsymbol{\psi}) = -\frac{1}{2} [n \lg(\sigma_Z^2) + \lg(\det \mathbf{R}(\boldsymbol{\psi})) + (\mathbf{Y}^n - \mathbf{F} \boldsymbol{\beta})^t \mathbf{R}(\boldsymbol{\psi})^{-1} (\mathbf{Y}^n - \mathbf{F} \boldsymbol{\beta}) / \sigma_Z^2]. \quad (2.23)$$

Dato $\boldsymbol{\psi}$, lo stimatore di massima verosimiglianza di $\boldsymbol{\beta}$ è dato da

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\psi}) = (\mathbf{F}^t \mathbf{R}(\boldsymbol{\psi})^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathbf{R}(\boldsymbol{\psi})^{-1} \mathbf{Y}^n \quad (2.24)$$

e lo stimatore di massima verosimiglianza di σ_Z^2

$$\hat{\sigma}_Z^2 = \hat{\sigma}_Z^2(\boldsymbol{\psi}) = \frac{1}{n}(\mathbf{Y}^n - \mathbf{F}\hat{\boldsymbol{\beta}}(\boldsymbol{\psi}))^t \mathbf{R}(\boldsymbol{\psi})^{-1}(\mathbf{Y}^n - \mathbf{F}\hat{\boldsymbol{\beta}}(\boldsymbol{\psi})) \quad (2.25)$$

Sostituendo (2.24) e (2.25) in (2.23), si ottiene:

$$l(\hat{\boldsymbol{\beta}}, \hat{\sigma}_Z^2, \boldsymbol{\psi}) = -\frac{1}{2}[n \lg \hat{\sigma}_Z^2(\boldsymbol{\psi}) + \lg(\det(\mathbf{R}(\boldsymbol{\psi}))) + n] \quad (2.26)$$

che dipende solamente da $\boldsymbol{\psi}$. Pertanto, lo stimatore di massima verosimiglianza è il valore di $\boldsymbol{\psi}$ che minimizza

$$n \lg \hat{\sigma}_Z^2(\boldsymbol{\psi}) + \lg(\det(\mathbf{R}(\boldsymbol{\psi}))),$$

con $\hat{\sigma}_Z^2$ definito da (2.25). Il previsore associato a $\hat{\boldsymbol{\psi}}$ viene chiamato MLE-EBLUP di Y_0 .

Massima Verosimiglianza Ristretta

Anche in questo caso si assume che \mathbf{R} e \mathbf{r}_0 dipendono da un vettore finito e non noto $\boldsymbol{\psi}$. Il metodo REML (Restricted Maximum Likelihood Estimation) è stato introdotto da Patterson e Thompson (1971) con lo scopo di determinare stimatori meno distorti rispetto a quelli di massima verosimiglianza. Lo stimatore REML di $\boldsymbol{\psi}$ massimizza la verosimiglianza di un insieme di combinazioni di \mathbf{Y}_n linearmente indipendenti, dove ogni combinazione lineare è ortogonale a $E(\mathbf{Y}_n) = \mathbf{F}\boldsymbol{\beta}$. Assumendo che \mathbf{F} ha rango pari a p , il metodo in questione consiste nello scegliere una matrice \mathbf{C} di dimensione $(n-p) \times n$ e rango $(n-p)$ che soddisfa $\mathbf{C}\mathbf{F} = \mathbf{0}$. Lo stimatore REML di $\boldsymbol{\psi}$ è quello che massimizza la verosimiglianza dei dati trasformati:

$$\mathbf{W} = \mathbf{C}\mathbf{Y}^n \sim N[\mathbf{C}\mathbf{F}\boldsymbol{\beta} = \mathbf{0}, \sigma_Z^2 \mathbf{C}\mathbf{R}(\boldsymbol{\psi})\mathbf{C}^t]$$

Osservazione 2.3 *Lo stimatore REML è indipendente dalla scelta delle combinazioni lineari usate per la costruzione di \mathbf{W} .*

Lo stimatore REML di σ_Z^2 è dato da

$$\tilde{\sigma}_Z^2 = \frac{n}{n-p} \hat{\sigma}_Z^2 = \frac{1}{n-p}(\mathbf{Y}^n - \mathbf{F}\hat{\boldsymbol{\beta}})^t \mathbf{R}(\boldsymbol{\psi})^{-1}(\mathbf{Y}^n - \mathbf{F}\hat{\boldsymbol{\beta}}), \quad (2.27)$$

con $\hat{\sigma}_Z^2$ lo stimatore MLE di σ_Z^2 (2.25) e lo stimatore REML di $\boldsymbol{\psi}$ è quello che minimizza

$$(n-p) \lg(\tilde{\sigma}_Z^2) + \lg(\det(\mathbf{R}(\boldsymbol{\psi}))) + \lg[\det(\mathbf{F}^t \mathbf{R}(\boldsymbol{\psi}) \mathbf{F})].$$

Altre metodologie sono state proposte in letteratura per la stima di $\boldsymbol{\psi}$ (ad esempio la cosiddetta *Cross-Validation*). Per approfondimenti sul argomento si veda Santner et al (2003).

2.4.3 Metodi Stocastici per Predizioni Spaziali (Kriging)

Il *Kriging* è un metodo di interpolazione che si usa per ottenere particolari previsori di processi stocastici o di funzioni aleatorie in punti in cui non sono state effettuate misurazioni. Infatti, nel campo della Statistica Spaziale, il Kriging è sinonimo di predizione spaziale. Esso deve il suo nome a Danie G. Krige, ingegnere minerario sud africano, pioniere nel campo della geostatistica. La tecnica *krigeage*, in francese, e *kriging*, in inglese, è stata denominata così in suo onore dall'ingegnere francese Georges Matheron (Matheron, 1962) che formalizzò il lavoro empirico di Krige. I metamodelli Kriging possono essere inquadrati nella classe dei Random Field Regression Models; infatti essi hanno la seguente forma

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x}),$$

dove $Z(\mathbf{x})$ è un processo stocastico Gaussiano stazionario di media $E[Z(\mathbf{x})] = 0$ e varianza $V[Z(\mathbf{x})] = \sigma_Z^2$ e $\mu(\cdot)$ un modello parametrico che specifica la struttura della media del processo stocastico $Y(\cdot)$. Ci sono diversi modelli Kriging tra i quali:

- *Kriging Semplice*: viene caratterizzato dal fatto che il trend globale è noto;
- *Kriging Ordinario*: il trend è costante, ossia $\mu(\mathbf{x}) = \mu$ ma ignoto. In questo caso il previsore di Y_0 (2.21) assume la forma:

$$\hat{Y}_0 = \hat{\beta} + \hat{\mathbf{r}}_0^t \hat{\mathbf{R}}^{-1} (\mathbf{Y}^n - \mathbf{1}\hat{\beta}) \quad (2.28)$$

- *Kriging Universale*: la componente di trend viene modellata in forma regressiva

$$\mu(\mathbf{x}) = \sum_{j=1}^p f_j(\mathbf{x})\beta_j = \mathbf{f}(\mathbf{x})^t \boldsymbol{\beta}.$$

Osservazione 2.4 *A questo punto è utile riflettere sulle conseguenze che comporta la scelta di $\mu(\mathbf{x})$ riguardanti la fedeltà del surrogato. Le Figure (2.1) e (2.2) mostrano gli effetti che ha l'ordine del polinomio del trend globale nell'adattamento del metamodello ai dati. In particolare, la Figura (2.1) mette in evidenza il fatto che la scelta di un trend globale costante (linea continua), lineare (linea tratteggiata) e quadratico (linea punteggiata) nel caso di **interpolazione**, comporta differenze trascurabili riguardanti l'adattamento dei tre metamodelli ai dati. Tuttavia, si notano differenze drastiche (si veda a riguardo la Figura (2.2)) nel caso di **estrapolazione**, ovvero quando le previsioni vengono fatte lontano dai dati a disposizione il modello Kriging*

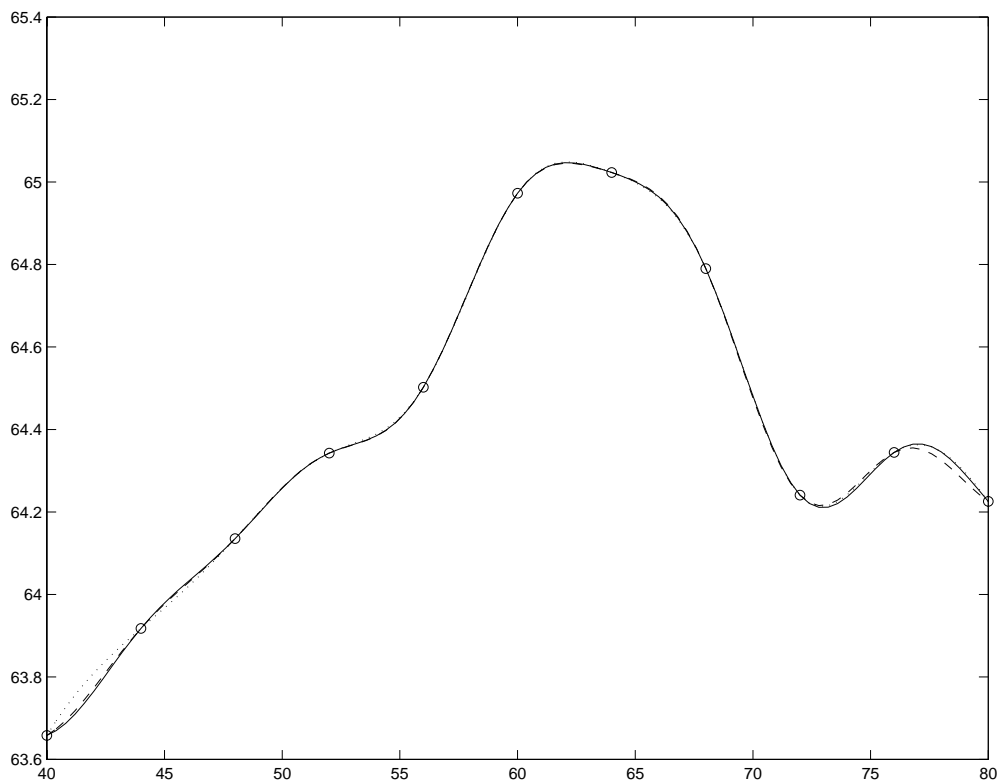


Figura 2.1: Illustrazione dell'impatto del trend nel caso di interpolazione.

○: dati ottenuti dal simulatore, linea continua: Kriging Ordinario, linea tratteggiata: Kriging Universale con trend lineare, linea punteggiata: Kriging Universale con trend quadratico.

degenera nel trend globale dal momento che si annulla il secondo termine della sommatoria in (2.21) e per questo motivo la scelta di esso deve essere fatta con molta cautela. E' prassi, comunque, evitare di usare il Kriging nel caso di estrapolazione. Per ulteriori dettagli si veda Sasena (2002), Kleijnen e van Beers (2004).

Osservazione 2.5 *In questo capitolo abbiamo illustrato solo alcuni dei modelli dei computer experiments. Altri tipi di surrogati, ad esempio le reti neurali o interpolatori Bayesiani, usati nell'ambito dei CE si possono trovare in Fang et al (2006).*

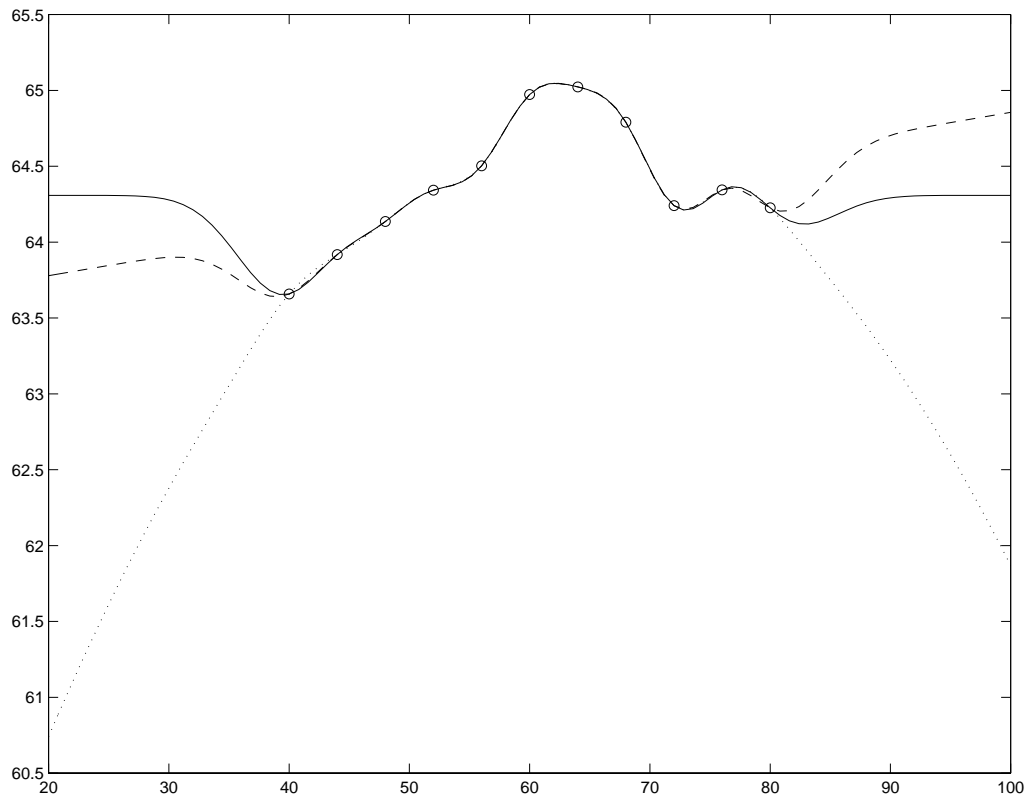


Figura 2.2: Illustrazione dell'impatto del trend nel caso di estrapolazione.

○: dati ottenuti dal simulatore, linea continua: Kriging Ordinario, linea tratteggiata: Kriging Universale con trend lineare, linea punteggiata: Kriging Universale con trend quadratico.

Capitolo 3

Esperimenti per la Stima del Metamodello: Disegni Space-Filling

3.1 Premessa

Abbiamo sottolineato più volte la differenza fondamentale tra la sperimentazione fisica e quella simulata. Nell'ambito dei computer experiments l'incertezza non nasce dall'errore casuale ma dal fatto che in molti casi il modello di simulazione non è esprimibile sotto una forma analitica trattabile o non si ha la conoscenza funzionale della relazione tra inputs e outputs, anche se la risposta può essere calcolata per ogni valore assunto dalla variabile di input. La discrepanza, quindi, tra l'output prodotto dal modello di simulazione e la previsione avuta dal metamodello scelto è l'errore da considerare. Se si ritiene che il modello vero presenti delle caratteristiche interessanti a più regioni del dominio sperimentale e che non ci siano delle regioni privilegiate nasce l'esigenza di osservare la risposta lungo l'intero spazio di disegno. I disegni Space-Filling, motivati proprio da questa considerazione, sono spesso basati su metodi di campionamento (ma anche su altri criteri come quello della distanza) e vengono chiamati in questo modo proprio per il fatto che mirano a distribuire i punti del dominio in modo equispaziato.

In letteratura ci sono varie tecniche che si possono adottare per realizzare una distribuzione in maniera uniforme dei punti di disegno. In questo capitolo approfondiremo tre approcci diversi motivati da questa esigenza: disegni basati su metodi di campionamento

(campionamento casuale semplice, campionamento stratificato, campionamento da ipercubo latino), quelli basati sul criterio della distanza e disegni uniformi. Per la presentazione dei disegni Space-Filling seguiamo Santner et al (2003) e Fang et al (2006).

3.2 Disegni basati su Metodi di Campionamento

3.2.1 Campionamento Casuale Semplice e Stratificato

La strategia più intuitiva per “riempire” lo spazio di disegno è quella di selezionare i punti in base a una griglia regolare. Ad esempio, supponendo che lo spazio di disegno sia il quadrato unitario $[0, 1] \times [0, 1]$ e che lo sperimentatore voglia osservare la risposta in 25 punti equispaziati, si potrebbe pensare di considerare la seguente griglia di punti

$$\{0.1, 0.3, 0.5, 0.7, 0.9\} \times \{0.1, 0.3, 0.5, 0.7, 0.9\}.$$

La metodologia statistica, e non solo intuitiva, più semplice che può essere usata per distribuire i punti in modo equispaziato è il campionamento casuale semplice, ovvero la selezione di un campione casuale semplice dallo spazio di disegno. Da un punto di vista teorico, il fatto che ci sono infiniti punti tra 0 e 1, rende tale scelta problematica. In pratica, però, è prassi considerare solo numeri con un numero finito di cifre decimali per cui il numero di punti compreso tra 0 e 1 può essere riguardato come finito, approssimazione che permette l'applicazione della tecnica statistica in questione. Tuttavia, nonostante il campionamento casuale semplice è utile in molte situazioni, può creare dei problemi in altre. Ad esempio, nel caso in cui la dimensione campionaria è piccola e lo spazio di disegno multi-dimensionale, il campione spesso presenta delle strutture di gruppo (clustering) per cui non riesce a fornire nessun punto in grandi regioni dello spazio di disegno. In questi casi, si usa il campionamento casuale stratificato: si selezionano n punti dividendo il dominio in n strati disposti uniformemente in tutto lo spazio di disegno e scegliendo in maniera casuale un punto da ogni strato. La possibilità di variare la dimensione e la posizione di ogni singolo strato e campionare rispetto a diverse distribuzioni rende flessibile la scelta del disegno in base allo scopo dell'esperimento.

Se si ha la convinzione del fatto che l'output dipende da pochi fattori di input (chiamato problema della sparsità), si vuole essere certi che i punti di disegno siano uniformemente distribuiti attraverso la proiezione dello spazio di disegno su tali fattori. Il campionamento da ipercubo latino (LHS: Latin Hypercube Sampling), che verrà presentato nel paragrafo

seguinte, rappresenta una metodologia valida da adottare per generare disegni i quali distribuiscono in modo equispaziato i punti sul range di ogni singolo fattore di input.

3.2.2 Campionamento da Ipercubi Latini

Questo paragrafo è dedicato alla descrizione del campionamento da ipercubi latini e le varie modifiche/sviluppi di esso. Si consideri il modello di simulazione

$$y = f(\mathbf{x}), \quad \mathbf{x} = (x_1, \dots, x_s)^t \in \mathcal{X}.$$

La tecnica legata agli ipercubi latini non si basa su una funzione specifica per il metamodello ma si focalizza sullo spazio di disegno che, senza perdita di generalità, spesso assumeremo pari a $\mathcal{C}^s = [0, 1]^s$. Partendo da qualsiasi dominio, con un'opportuna riparametrizzazione si può ottenere \mathcal{C}^s .

Definizione 3.1 Un *disegno da ipercubo latino* costituito da n prove (runs) e s variabili di input è una matrice $n \times s$ tale che ogni sua colonna è una permutazione casuale di $\{1, 2, \dots, n\}$ e viene denotato con $\text{LHD}(n, s)$:

Un $\text{LHS}(n, s)$ può essere fornito dall'algoritmo di cui sotto:

Passo I Generare in modo indipendente s permutazioni $\pi_j(1), \dots, \pi_j(n)$ dei numeri interi $1, \dots, n$ con $j = 1, \dots, s$, ovvero generare un $\text{LHD}(n, s)$;

Passo II Prendere $n \cdot s$ numeri random $U_k^j \sim U(0, 1)$, dove $k = 1, \dots, n$ e $j = 1, \dots, s$. Siano $\mathbf{x}_k = (x_k^1, \dots, x_k^s)$ dove

$$x_k^j = \frac{\pi_j(k) - U_k^j}{n}, \quad k = 1, \dots, n, \quad j = 1, \dots, s$$

Allora il disegno $\mathcal{D}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ è un $\text{LHS}(n, s)$.

Il disegno precedentemente presentato è stato proposto da McKay et al (1979) e viene considerato da molti il primo lavoro sul disegno degli esperimenti finalizzato ai computer experiments.

Esempio 3.2 Il primo passo per generare un LHS con $n = 8$ e $s = 2$ consiste nel generare due permutazioni di $\{1, 2, \dots, 8\}$. Successivamente, si generano $8 \cdot 2$ numeri casuali tra 0 e

1 per formare una matrice 8×2 , ossia

$$\begin{bmatrix} 2 & 5 \\ 5 & 8 \\ 1 & 3 \\ 7 & 6 \\ 4 & 1 \\ 8 & 4 \\ 3 & 7 \\ 6 & 2 \end{bmatrix}, \begin{bmatrix} 0.9501 & 0.8214 \\ 0.2311 & 0.4447 \\ 0.6068 & 0.6154 \\ 0.4860 & 0.7919 \\ 0.8913 & 0.9218 \\ 0.7621 & 0.7382 \\ 0.4565 & 0.1763 \\ 0.0185 & 0.4057 \end{bmatrix}$$

Di conseguenza, il disegno da ipercubo latino è dato da:

$$\frac{1}{8} \begin{bmatrix} \begin{bmatrix} 2 & 5 \\ 5 & 8 \\ 1 & 3 \\ 7 & 6 \\ 4 & 1 \\ 8 & 4 \\ 3 & 7 \\ 6 & 2 \end{bmatrix} - \begin{bmatrix} 0.9501 & 0.8214 \\ 0.2311 & 0.4447 \\ 0.6068 & 0.6154 \\ 0.4860 & 0.7919 \\ 0.8913 & 0.9218 \\ 0.7621 & 0.7382 \\ 0.4565 & 0.1763 \\ 0.0185 & 0.4057 \end{bmatrix} = \begin{bmatrix} 0.1312 & 0.5223 \\ 0.5961 & 0.9444 \\ 0.0491 & 0.2981 \\ 0.8143 & 0.6510 \\ 0.3886 & 0.0098 \\ 0.9047 & 0.4077 \\ 0.3179 & 0.8530 \\ 0.7477 & 0.1993 \end{bmatrix}$$

Si noti che $(\pi_1(k), \pi_2(k))$ indicano in quale cella viene posto il punto di disegno \mathbf{x}_k , mentre (U_k^1, U_k^2) determinano dove deve essere posizionato \mathbf{x}_k all'interno della cella.

Al fine di ridurre la complessità del campionamento, si potrebbe sfruttare per la costruzione del LHS una struttura a griglia al posto dei numeri random U_k^j , ovvero

$$x_k^j = \frac{\pi_j(k) - 0.5}{n}, \quad k = 1, \dots, n, \quad j = 1, \dots, s$$

equivale a dire, posizionare il punto di disegno \mathbf{x}_k al centro di ogni cella. Il LHS ottenuto in questo modo viene chiamato “midpoint Latin Hypercube Sampling” oppure “centered Latin Hypercube Sampling” e viene indicato MLHS(n, s).

Il cosiddetto “U-type design” (Fang e Hickernell, 1995) è un’ulteriore disegno volto a distribuire i punti del dominio in modo equispaziato:

Definizione 3.3 Un *U-type design* costituito da n prove e s fattori, ognuno avente q_1, \dots, q_s livelli, rispettivamente, è una matrice $n \times s$ tale che i q_j livelli si presentano nella j -esima

colonna lo stesso numero di volte. Questo disegno viene denotato con $U(n, q_1 \times \dots \times q_s)$. Nel caso in cui alcuni dei livelli q_j coincidono, il disegno in questione si denota con $U(n, q_1^{r_1} \times \dots \times q_m^{r_m})$, con r_1, \dots, r_m numeri interi tali che $r_1 + \dots + r_m = s$. Se tutti i livelli q_j coincidono, allora il disegno si denota con $U(n, q^s)$ e viene detto simmetrico, altrimenti viene chiamato asimmetrico.

Un esempio di un disegno $U(9, 3^4)$ è dato da:

Esempio 3.4

$\mathbf{U}(9, \mathbf{3}^4)$	1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

In genere, tutti i disegni LHS

- (i) risultano facili da generare da un punto di vista computazionale;
- (ii) possono essere usati anche quando il numero di prove e il numero di variabili di input è grande.

3.3 Disegni basati sulla Distanza

In questo paragrafo prenderemo in considerazione dei criteri per la selezione di un disegno basati su una specifica misura o metrica (Johnson et al, 1990). Essi vengono caratterizzati dal fatto che misurano quanto uniformemente sono distribuiti i punti presi in esame. Un approccio per realizzare tale distribuzione consiste nell'imporre la "non troppa" vicinanza tra due punti. In particolare, siano $\mathcal{D} \subset \mathcal{X} \subset \mathbb{R}^s$ un disegno costituito da n punti distinti $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ e ρ una metrica su \mathcal{X} definita, per $p \geq 1$, da

$$\rho_p(\mathbf{w}, \mathbf{x}) = \left(\sum_{j=1}^s |w_j - x_j|^p \right)^{\frac{1}{p}}$$

Una misura di vicinanza dei punti di \mathcal{D} è la minima distanza tra due punti qualsiasi di \mathcal{D} , ovvero

$$\min_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}} \rho_p(\mathbf{x}_1, \mathbf{x}_2).$$

Un disegno che massimizza tale misura viene chiamato *disegno di distanza maximin* e denotato con \mathcal{D}_{Mm} . Quindi si ha

$$\min_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}_{Mm}} \rho_p(\mathbf{x}_1, \mathbf{x}_2) = \max_{\mathcal{D} \subset \mathcal{X}} \min_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}} \rho_p(\mathbf{x}_1, \mathbf{x}_2)$$

Un modo alternativo per realizzare una distribuzione dei punti di disegno più uniforme possibile in \mathcal{X} è quello di richiedere che ogni punto dello spazio \mathcal{X} sia “vicino” a qualche punto appartenente al disegno \mathcal{D} . In particolare, si definisce la distanza tra qualsiasi $\mathbf{x} \in \mathcal{X}$ e il disegno $\mathcal{D} \subset \mathcal{X}$ nel seguente modo

$$\rho_p(\mathbf{x}, \mathcal{D}) = \min_{\mathbf{x}_i \in \mathcal{D}} \rho_p(\mathbf{x}, \mathbf{x}_i). \quad (3.1)$$

Un disegno \mathcal{D} costituito da n punti viene detto *disegno di distanza minimax* e denotato con \mathcal{D}_{mM} se la distanza massima tra un punto qualsiasi $\mathbf{x} \in \mathcal{X}$ e il disegno in questione risulta minima nella classe di tutti i disegni $\mathcal{D} \subset \mathcal{X}$, ossia

$$\min_{\mathcal{D} \subset \mathcal{X}} \max_{\mathbf{x} \in \mathcal{X}} \rho_p(\mathbf{x}, \mathcal{D}) = \max_{\mathbf{x} \in \mathcal{X}} \rho_p(\mathbf{x}, \mathcal{D}_{mM}). \quad (3.2)$$

3.4 Disegni Uniformi

Nei paragrafi precedenti abbiamo esaminato i disegni Space-Filling legati a vari metodi di campionamento e quelli basati sul criterio della distanza tra punti. Il terzo principio si basa, invece, sul confronto tra la distribuzione dei punti di disegno e la distribuzione uniforme.

Supponiamo anche in questo contesto che il vettore degli input $\mathbf{x} = (x_1, \dots, x_s)$ sia s -dimensionale e che lo spazio di disegno sia $\times_{i=1}^s [a_i, b_i]$. Inoltre, sia $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ l'insieme degli n punti di disegno. Useremo la notazione \mathbf{X} al posto di \mathbf{x} nei casi in cui vorremo sottolineare il fatto che essa sia una variabile casuale. Infine, la distribuzione uniforme della variabile casuale \mathbf{X} verrà denotata come segue

$$\mathbf{X} \sim F(\cdot), \quad \text{dove} \quad F(\mathbf{x}) = \prod_{i=1}^s \left(\frac{x_i - a_i}{b_i - a_i} \right)$$

Fang et al (2000) hanno illustrato e commentato la nozione di “discrepanza” di un dato disegno \mathcal{D} , che misura l'entità con la quale \mathcal{D} differisce dalla distribuzione uniforme dei

punti. In particolare, sia F_n la funzione di distribuzione empirica dei punti di \mathcal{D} , ovvero

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n I\{\mathbf{X}_i \leq \mathbf{x}\}, \quad (3.3)$$

dove $I\{E\}$ è la funzione indicatrice di E e la disuguaglianza in (3.3) si fa rispetto all'ordinamento per componenti corrispondenti (componentwise ordering) di vettori in \mathbb{R}^s . La L_∞ -discrepanza verrà indicata con $D_\infty(\mathcal{D})$ e definita come segue

$$D_\infty(\mathcal{D}) = \sup_{\mathbf{x} \in \mathcal{X}} |F_n(\mathbf{x}) - F(\mathbf{x})|.$$

La misura di discrepanza di cui sopra è una delle più popolari in letteratura. Un'altra misura di discrepanza molto usata è la L_p la quale viene denotata con $D_p(\mathcal{D})$ e data da

$$D_p(\mathcal{D}) = \left[\int_{\mathcal{X}} |F_n(\mathbf{x}) - F(\mathbf{x})|^p d\mathbf{x} \right]^{\frac{1}{p}}$$

Si noti che la L_∞ -discrepanza è un caso limite della L_p -discrepanza per $p \rightarrow \infty$. Disegni costituiti da un insieme di punti con discrepanza piccola vengono considerati più uniformi rispetto a quelli con grande discrepanza. Inoltre, i disegni costituiti da punti che minimizzano D_p vengono chiamati *disegni uniformi*.

3.5 Disegni basati su Combinazioni di Criteri

Finora abbiamo esaminato vari criteri che possono essere usati per generare disegni Space-Filling: criteri basati su metodi di campionamento, disegni Space-Filling generati da criteri di distanza ed in particolare disegni maximin e minimax, e disegni uniformi. Purtroppo, nessuno dei criteri precedentemente analizzati gode di proprietà tali da renderlo soddisfacente in tutte le circostanze per essere usato come strumento di selezione di un disegno.

Un approccio per ovviare alla suddetta problematica, è quello di usare la combinazione di più criteri per generare disegni che godono di proprietà desiderate. In particolare, si potrebbe sfruttare, in primo luogo, un criterio per ridurre l'insieme dei punti-candidati come punti di disegno e, in seguito, applicare un secondo criterio per selezionare il disegno finale all'interno della classe ristretta, precedentemente presa in considerazione. Ad esempio, tra tutti i disegni LHS si potrebbe scegliere quello più uniforme. Sostanzialmente, tale procedimento coinvolgerebbe l'utilizzo, in una prima fase, del campionamento da ipercubo latino seguito dall'applicazione del criterio basato sulla discrepanza.

Nella letteratura ci sono numerosi articoli riguardanti disegni fondati sulla combinazione di vari criteri. Morris and Mitchell (1995) hanno generato disegni sfruttando il criterio di distanza maximin nella classe dei disegni LHS, Iman e Conover (1982) e Owen (1994) hanno proposto dei disegni nati dalla classe LHS applicando ad essa un criterio di correlazione lineare, mentre Tang (1998) ha modificato il lavoro degli ultimi applicando, sempre nella classe degli LHS, un criterio di correlazione canonica polinomiale. Una vasta gamma di lavori basati sulla combinazione di criteri possono essere trovati nella letteratura a parte quelli citati in precedenza.

3.6 Applicazione della Metodologia Kriging e dei Disegni Space-Filling

Questo paragrafo sarà dedicato alla presentazione di un caso di studio (Gupta et al, 2006) riguardante il campo della telefonia mobile.

L'identificazione del settaggio ideale, ossia la scelta dei valori delle variabili di ingresso che ottimizzano la risposta è un passaggio cruciale per il miglioramento della qualità dei processi. Tradizionalmente, uno degli approcci maggiormente utilizzati nell'ambito della produzione industriale è rappresentato dalla *Metodologia delle Superfici di Risposta* applicata ad esperimenti fisici. Al giorno d'oggi le simulazioni al computer, ad esempio le simulazioni basate sull'analisi agli elementi finiti (FEA: Finite Element Analysis), vengono usate per sostituire la sperimentazione fisica. Tuttavia, tali simulazioni sono solitamente costose in termini computazionali data la loro complessità e pertanto si ricorre all'uso dei metamodelli. In questo lavoro, attraverso la metodologia Kriging abbinata ad una procedura sequenziale, gli autori riescono ad individuare i valori ideali delle variabili di input che minimizzano la variabile di risposta con il 40% in meno di simulazioni rispetto all'approccio delle Superfici di Risposta.

3.6.1 Introduzione

Dato il mercato sempre più competitivo, c'è la necessità di ideare e costruire prodotti tecnicamente più complessi che siano in grado di garantire un'elevata affidabilità, il tutto in tempi assai più brevi di quelli richiesti dalla sperimentazione fisica. Nel mercato dei telefoni cellulari ciò si traduce in una maggiore versatilità, trasportabilità e attrazione visiva ed ergonomica, il che a sua volta richiede un miglioramento continuo nell'assem-

blaggio elettronico. Una problematica riscontrata in questo ambito consiste nel fallimento della lega delle giunture, ossia la lega che congiunge due parti del cellulare, quando viene sottoposta ad un carico termale e meccanico. Il livello dello stress di von Mises (relativo cioè al criterio di von Mises o dell'energia di deformazione) generato nelle giunture è considerato la caratteristica chiave per valutare l'affidabilità della lega e pertanto deve essere minimizzato (passo 1 della procedura di sviluppo dell'emulatore); tale livello viene quindi identificato come variabile di risposta (passo 2), in corrispondenza della quale un team di ingegneri esperti ha selezionato quattro variabili di input legate alle caratteristiche del materiale associato con la lega delle giunture (passo 3):

$$x_1 \in [20, 35], x_2 \in [12, 36], x_3 \in [1, 5] \text{ e } x_4 \in [15, 30].$$

In sintesi, l'obiettivo dello studio in questione è la costruzione di un opportuno emulatore per un problema di ottimizzazione.

3.6.2 Metodologia

Gli autori adottano come metamodello il Kriging Ordinario (passo 5) e assumono la seguente funzione di correlazione

$$\mathcal{R}(\|\mathbf{x} - \mathbf{x}'\|) = \prod_{j=1}^4 \exp(-\theta_j |x_j - x'_j|^2),$$

dove x_j, x'_j rappresentano la j -esima componente di \mathbf{x} e \mathbf{x}' , rispettivamente, e $\theta_j \geq 0$. Pertanto, i parametri non noti sono 6, ovvero

$$\sigma_Z^2, \quad \boldsymbol{\theta} \equiv (\theta_1, \theta_2, \theta_3, \theta_4) \quad \text{e} \quad \beta.$$

Per quanto concerne la pianificazione sperimentale (passo 6) gli autori propongono una procedura sequenziale volta ad identificare il settaggio dei fattori di input che minimizza il livello di stress nell'assemblaggio elettronico dei telefoni, e l'uso dei disegni ad ipercubi latini. Di seguito descriviamo passo per passo la procedura proposta.

1) Inizialmente è stato adottato un LHS(9, 4):

Prove	x_1	x_2	x_3	x_4	y
1	21.58	20.37	3.50	17.64	158.62
2	25.85	25.44	4.66	22.79	169.37
3	30.74	22.91	2.70	26.03	165.25
4	23.31	18.65	2.70	26.03	158.62
5	33.90	35.65	1.21	20.52	180.24
6	31.75	32.90	3.08	19.22	176.60
7	29.56	16.25	3.88	26.74	175.45
8	24.56	12.63	1.73	29.21	167.88
9	27.48	29.99	2.33	15.24	168.63

] ; (3.4)

2) in base alle osservazioni simulate sui punti del disegno (3.4) ottenuti al passo precedente sono stati stimati i parametri incogniti del Kriging Ordinario attraverso il metodo della massima verosimiglianza

$$\hat{\sigma}_Z^2 = 37.81, \hat{\boldsymbol{\theta}} = (2.5341, 0.6286, 0.4502, 1.0967) \quad \text{e} \quad \hat{\beta} = 170.47;$$

3) successivamente lo spazio di disegno è stato ristretto in una sotto-regione, ovvero è stata selezionata la regione del dominio che presentava più incertezza. In particolare, sono stati tracciati i grafici dell'errore quadratico medio avendo preso in considerazione le variabili a due a due (ad esempio x_1 e x_2 nella Figura (3.1)) e ognuno di essi ha identificato una restrizione del dominio per ciascuna variabile di input data dalla regione del grafico in cui l'errore quadratico medio è risultato più elevato (zona rossa della Figura (3.1) dove la sotto-regione individuata è data da $x_1 \in [0.8, 1]$ e $x_2 \in [0, 0.3]$ che nello spazio originale corrisponde a $x_1 \in [32, 35]$ e $x_2 \in [12, 19.2]$). Alla fine è stato preso come spazio di disegno aggiornato l'unione delle suddette regioni;

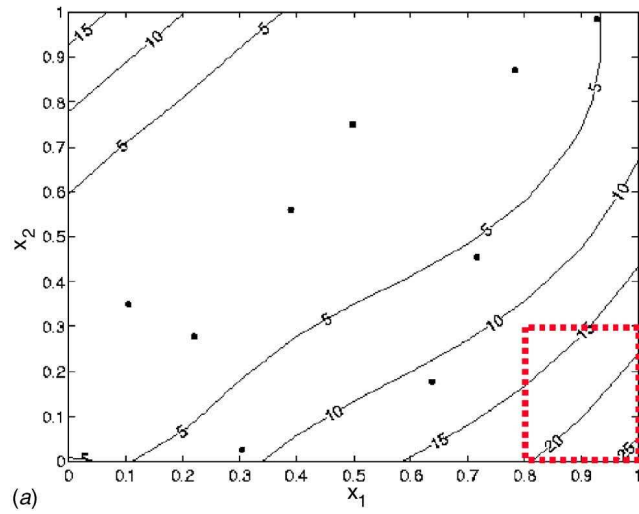


Figura 3.1: Grafico dell'errore quadratico medio di x_1 e x_2 .

4) è stato adottato di nuovo un LHS a partire dallo spazio ridotto individuato:

Prove	x_1	x_2	x_3	x_4	y
10	34.93	23.94	3.77	23.29	181.38
11	32.39	13.013	2.65	29.05	179.08
12	33.40	34.22	1.53	17.92	178.09
13	32.69	16.27	1.89	21.24	176.49
14	34.36	26.14	4.44	25.73	182.02
15	33.90	30.33	3.65	16.43	177.14

(3.5)

e in base alle osservazioni simulate sui punti di cui sopra sono state aggiornate le stime dei parametri: $\hat{\sigma}_Z^2 = 27.2614$, $\hat{\theta} = (4.1054, 0.5849, 0.3153, 0.8876)$ e $\hat{\beta} = 171.82$;

5) come il passo 2 con la differenza che la sotto-regione deve essere una regione del dominio privilegiata che possa contenere il settaggio ideale che minimizza l'output; essa è stata scelta attraverso i plot di iterazione tra i fattori di input presi a due a due. Siccome il minimo valore della risposta che è stato fornito dalle prime quindici simulazioni è risultato pari a 158.62 (si veda (3.4) e (3.5)), valore che indica orientativamente le zone dove cercare il settaggio ideale che minimizza la variabile di risposta, in questo step gli autori hanno ristretto ancora di più lo spazio di disegno in base al valore presunto minimo di y (che hanno assunto prossimo a 158) nei grafici

di iterazione (zona rossa della Figura (3.2) dove la sotto-regione individuata è data da $x_1 \in [0, 0.1]$ e $x_2 \in [0, 0.5]$);

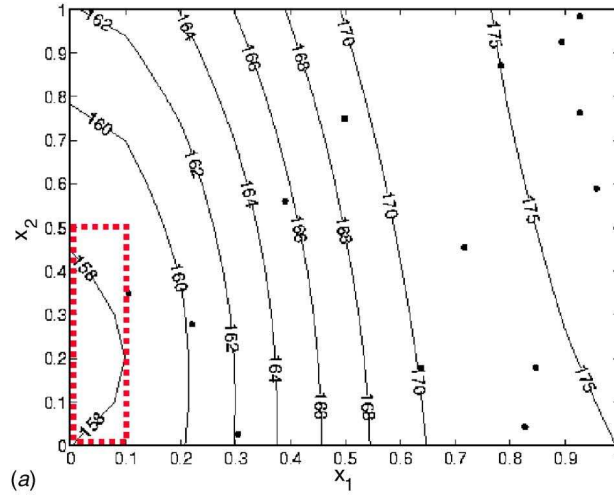


Figura 3.2: Plot di iterazione tra x_1 e x_2 basato sul metamodella previsivo del quarto step.

6) a partire dal nuovo spazio ridotto è stato adoperato un ulteriore disegno LHS(6, 4) dato da

Prove	x_1	x_2	x_3	x_4	y
16	21.47	17.97	3.22	19.15	158.78
17	20.20	12.51	2.32	22.03	156.82
18	20.70	23.11	1.43	16.46	156.61
19	20.35	14.14	1.71	18.12	155.59
20	21.18	19.07	3.75	20.36	158.89
21	20.95	21.16	3.12	15.71	156.35

e sono stati stimati i parametri ottenendo:

$$\hat{\sigma}_Z^2 = 56.89, \hat{\theta} = (2.2029, 0.1449, 0.8688, 4.7316) \quad \text{e} \quad \hat{\beta} = 167.83;$$

7) la procedura di selezione delle sotto-regioni privilegiate è stata fermata quando è stata ottenuta l'accuratezza del metamodella desiderata, ovvero quando gli incrementi/decrementi dei valori di minimo della risposta non risultavano significativi e i plot dell'errore quadratico medio non presentavano valori elevati d'incertezza.

Capitolo 4

Disegni Ottimi per il Kriging Ordinario

Nel capitolo precedente abbiamo esaminato una ampia classe di procedure le quali nascono dall'esigenza di osservare la risposta lungo l'intero spazio di disegno e mirano a distribuire i punti del dominio in modo equispaziato. Un metodo alternativo riguardante la selezione dei valori per le variabili di input nell'ambito dei computer experiments per la stima dei parametri dei metamodelli parametrici viene, invece, basato sulla teoria dei disegni ottimi e sarà adottato qui.

L'argomento di questo capitolo è la ricerca di piani di esperimento per la stima di massima verosimiglianza dei parametri del Kriging Ordinario unidimensionale avente struttura di correlazione esponenziale parametrizzata. Fissato quindi lo scopo dell'esperimento, dal punto di vista della pianificazione il problema consiste nello scegliere il disegno che minimizza la matrice di varianza covarianza degli stimatori di massima verosimiglianza dei parametri del modello.

4.1 Premessa

Si consideri un contesto generale nel quale si vuole osservare un processo continuo Y in un insieme discreto di punti i quali possono essere scelti dallo sperimentatore in uno spazio di disegno compatto \mathcal{X} e sia $y(\mathbf{x}) = f(x_1, \dots, x_s)$ un'osservazione in un dato punto \mathbf{x} la quale, come abbiamo già detto nel secondo capitolo, può essere vista come una realizzazione di

un processo stocastico

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x}), \quad \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^s$$

con $Z(\mathbf{x})$ un processo Gaussiano stazionario di media zero, varianza costante σ_Z^2 e funzione di correlazione tra due input \mathbf{x} e \mathbf{x}' data da

$$\mathcal{R}(Z(\mathbf{x}), Z(\mathbf{x}')) = \mathcal{R}(\|\mathbf{x} - \mathbf{x}'\|)$$

la quale tende a uno per $\|\mathbf{x} - \mathbf{x}'\| \rightarrow 0$.

Diversi approcci inferenziali sono stati usati in letteratura per la stima dei parametri della funzione di correlazione. Nell'ambito della statistica spaziale, spesso vengono adottati approcci non parametrici basati sul variogramma empirico (Matheron (1962), Cressie e Hawkins (1980), Bogaert e Russo (1999), Müller e Zimmerman (1999)).

D'altra parte, in molti contesti applicativi (ad esempio computer experiments e geostatistica) succede spesso che lo sperimentatore possieda dell'informazione riguardante le proprietà di secondo ordine del processo $Y(\cdot)$. In questi casi, si adopera un approccio inferenziale il quale si basa su una specifica famiglia parametrica per la funzione di covarianza (esponenziale, Gaussiana, Matérn...), le osservazioni vengono usate per stimare i parametri d'interesse del modello e la predizione di Y viene basata sulla stima della media e della funzione di covarianza del processo.

Recentemente, molti autori (Abt e Welch (1998), Park e Baek (2001), Pázman (2004), Li e Sudjianto (2005), Zhang e Zimmermann (2005), Zhu e Stein (2005), Zimmermann (2006)) hanno adottato procedure inferenziali incentrate sulla funzione di verosimiglianza (ad esempio, massima verosimiglianza, massima verosimiglianza ristretta, verosimiglianza penalizzata e verosimiglianza profilo). In particolare, Pázman (2004) ha dimostrato che la matrice di varianza covarianza degli stimatori di massima verosimiglianza dei parametri può essere approssimata attraverso l'inversa della matrice di informazione di Fisher (FIM). Come mostrato da Zhu e Stein (2005) tramite uno studio di simulazioni, tale approssimazione è buona nel caso in cui la dimensione campionaria è abbastanza elevata; in caso contrario, essa non è accurata ma la matrice di informazione di Fisher può ancora essere adottata come criterio di disegno.

4.2 Modello

Il modello, oggetto di studio di questo capitolo, è il Kriging Ordinario unidimensionale

$$Y(x) = \mu + Z(x), \quad x \in \mathcal{X} \subset \mathbb{R} \quad (4.1)$$

con le usuali ipotesi per $Z(\cdot)$, avente struttura di covarianza $Cov(Z(x), Z(x'))$ dipendente solo dalla distanza tra x e x' data da

$$Cov(Z(x), Z(x')) = \sigma_Z^2 \exp\{-\theta|x - x'|\}, \quad \theta \in \mathbb{R}^+. \quad (4.2)$$

Questo particolare processo viene spesso chiamato processo di Ornstein-Uhlenbeck. Si noti che il modello (4.1) rientra nella classe dei modelli parzialmente lineari dato che il valore atteso della risposta Y è lineare rispetto a μ .

La funzione di correlazione dipende da un parametro positivo θ il quale esprime l'intensità della dipendenza spaziale: valori piccoli di θ rispecchiano una correlazione grande tra coppie di osservazioni mentre al crescere di θ le osservazioni prese in punti anche vicini tendono ad essere incorrelate. Inoltre, le osservazioni fatte al medesimo punto hanno correlazione unitaria (ricordiamo che i computer experiments sono deterministici) per cui le repliche non danno informazione aggiuntiva, motivo per cui di seguito non prenderemo in considerazione disegni con osservazioni replicate allo stesso punto.

Supponiamo adesso di condurre un esperimento scegliendo n punti, x_i ($i = 1, \dots, n$), e denotiamo con $\xi_n = \{x_1, \dots, x_n\}$ il disegno, ossia l'insieme di questi punti nei quali si osserverà il processo $Y(\cdot)$. Stehlík (2007) studia i disegni ottimi per le stime separate dei parametri presenti nel modello Kriging Ordinario quando il disegno consiste in pochi punti. Inoltre, Pepelyshev (2007) trova disegni D-ottimi, ossia disegni che massimizzano il determinante della matrice di informazione di Fisher, per il modello di regressione esponenziale con osservazioni correlate e struttura di errori autoregressiva considerando un numero limitato di punti.

Senza perdita di generalità consideriamo $\mathcal{X} = [0, 1]$ e

$$0 \leq x_1 < x_2 < \dots < x_{n-1} \leq 1.$$

Denotando con $\mathbf{Y} = (Y(x_1), \dots, Y(x_n))^t$ il vettore che rappresenta le osservazioni fatte nei punti di disegno, il modello può essere riscritto nel seguente modo

$$\mathbf{Y} \sim N(\boldsymbol{\mu}; \sigma_Z^2 \mathbf{R}) \quad (4.3)$$

dove $\boldsymbol{\mu} = (\mu, \dots, \mu)^t$ è il vettore $(n \times 1)$ -dimensionale della componente trend del modello e \mathbf{R} è la matrice di correlazione data da

$$\mathbf{R} = \begin{pmatrix} 1 & \exp(-\theta(x_2 - x_1)) & \dots & \exp(-\theta(x_n - x_1)) \\ \exp(-\theta(x_2 - x_1)) & 1 & \dots & \exp(-\theta(x_n - x_2)) \\ \vdots & \vdots & \ddots & \vdots \\ \exp(-\theta(x_n - x_1)) & \exp(-\theta(x_n - x_2)) & \dots & 1 \end{pmatrix} \quad (4.4)$$

L'interesse di questo capitolo è incentrato sui parametri μ e θ del modello di cui sopra, mentre σ_Z^2 viene considerato noto. Essendo il Kriging un estrapolatore pessimo (si veda Osservazione 2.4), prenderemo in considerazione solo piani di esperimento che mettono due dei punti di disegno agli estremi di \mathcal{X} , i.e. $x_1 = 0$ e $x_n = 1$. Il problema allora si traduce nello scegliere un vettore di $n - 2$ punti di disegno ($n > 2$) $x_i \in (0, 1)$ ($i = 2, \dots, n - 1$) per la stima di massima verosimiglianza dei parametri.

4.3 La Matrice di Informazione di Fisher

La log verosimiglianza del nostro modello (4.1) è data da

$$l(\mu, \theta; \xi_n) = -\frac{1}{2} \left\{ \frac{(\mathbf{Y} - \boldsymbol{\mu})^t \mathbf{R}^{-1} (\mathbf{Y} - \boldsymbol{\mu})}{\sigma_Z^2} + \lg \det \mathbf{R} + n \lg(2\pi\sigma_Z^2) \right\} \quad (4.5)$$

per cui la matrice d'informazione di Fisher può essere scritta nel seguente modo

$$\mathbf{I}(\mu, \theta; \xi_n) = \begin{pmatrix} \frac{\mathbf{1}^t \mathbf{R}^{-1} \mathbf{1}}{\sigma_Z^2} & 0 \\ 0 & \frac{1}{2} \text{tr} \left[(\mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \theta})^2 \right] \end{pmatrix} \quad (4.6)$$

Denotando con $d_i = x_{i+1} - x_i$ (per $i = 1, \dots, n - 1$) la distanza tra due punti adiacenti, la matrice di correlazione (4.4) può essere riscritta come segue:

$$\mathbf{R} = \begin{pmatrix} 1 & \exp(-\theta d_1) & \dots & \exp(-\theta \sum_{i=1}^{n-1} d_i) \\ \exp(-\theta d_1) & 1 & \dots & \exp(-\theta \sum_{i=2}^{n-1} d_i) \\ \vdots & \vdots & \ddots & \vdots \\ \exp(-\theta \sum_{i=1}^{n-1} d_i) & \exp(-\theta \sum_{i=2}^{n-1} d_i) & \dots & 1 \end{pmatrix}, \quad (4.7)$$

Siccome non vengono presi in considerazione disegni con osservazioni replicate allo stesso punto e dall'assunzione $x_1 = 0$ e $x_n = 1$, si ha che

$$\sum_{i=1}^{n-1} d_i = 1 \quad \text{con} \quad d_i \in (0; 1) \quad \text{for} \quad i = 1, \dots, n - 1 \quad (4.8)$$

per cui la matrice \mathbf{R} è non singolare. Si noti che ogni disegno composto da n punti genera univocamente $n - 1$ distanze e viceversa. Pertanto, se indichiamo con $\mathbf{d}_{\xi_n} = (d_1, \dots, d_{n-1})$ il vettore delle distanze associate a ξ_n , da (4.6) e (4.7) si ottiene che la matrice di informazione di Fisher $\mathbf{I}(\mu, \theta; \xi_n)$ dipende dal disegno solo attraverso \mathbf{d}_{ξ_n} e quindi

$$\mathbf{I}(\mu, \theta; \xi_n) = \mathbf{I}(\mu, \theta; \mathbf{d}_{\xi_n}) = \begin{pmatrix} I_\mu(\mathbf{d}_{\xi_n}) & 0 \\ 0 & I_\theta(\mathbf{d}_{\xi_n}) \end{pmatrix}, \quad (4.9)$$

dove $I_\mu(\mathbf{d}_{\xi_n})$ e $I_\theta(\mathbf{d}_{\xi_n})$ sono le informazioni di μ e θ , rispettivamente.

Teorema 4.1 *L'informazione $I_\mu(\mathbf{d}_{\xi_n})$ del trend μ può essere scritta nel seguente modo*

$$I_\mu(\mathbf{d}_{\xi_n}) = \frac{1}{\sigma_Z^2} \left[1 + \sum_{i=1}^{n-1} f(d_i) \right], \quad \text{dove} \quad f(y) = \frac{\exp(\theta y) - 1}{\exp(\theta y) + 1}.$$

Dimostrazione 4.2 *Sfruttando la decomposizione **LDU** per matrici simmetriche (si veda a riguardo Harville, 1997), si ottiene per la matrice di correlazione **R**:*

$$\mathbf{R} = \mathbf{L}\mathbf{D}\mathbf{L}^t \quad (4.10)$$

dove \mathbf{D} è una matrice diagonale

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 - \exp(-2\theta d_1) & 0 & \dots & 0 \\ 0 & 0 & 1 - \exp(-2\theta d_2) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 - \exp(-2\theta d_{n-1}) \end{pmatrix} \quad (4.11)$$

e \mathbf{L} è una matrice triangolare inferiore data da

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \exp(-\theta d_1) & 1 & 0 & \dots & 0 \\ \exp(-\theta \sum_{i=1}^2 d_i) & \exp(-\theta d_2) & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \exp(-\theta \sum_{i=1}^{n-1} d_i) & \exp(-\theta \sum_{i=2}^{n-1} d_i) & \exp(-\theta \sum_{i=3}^{n-1} d_i) & \dots & 1 \end{pmatrix}. \quad (4.12)$$

Da (4.10) si ha che

$$\mathbf{R}^{-1} = (\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1})^t(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}),$$

dove

$$\mathbf{L}^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\exp(-\theta d_1) & 1 & 0 & \dots & 0 \\ 0 & -\exp(-\theta d_2) & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -\exp(-\theta d_{n-1}) & 1 \end{pmatrix},$$

per cui

$$I_\mu(\mathbf{d}_{\xi_n}) = \frac{1}{\sigma_Z^2} \mathbf{1}^t \mathbf{R}^{-1} \mathbf{1} = \frac{1}{\sigma_Z^2} \left[(\mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \mathbf{1})^t (\mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \mathbf{1}) \right]$$

dove

$$\mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \mathbf{1} = \left(1, \frac{1 - \exp(-\theta d_1)}{\sqrt{1 - \exp(-2\theta d_1)}}, \dots, \frac{1 - \exp(-\theta d_{n-1})}{\sqrt{1 - \exp(-2\theta d_{n-1})}} \right)^t.$$

Pertanto

$$I_\mu(\mathbf{d}_{\xi_n}) = \frac{1}{\sigma_Z^2} \left[1 + \sum_{i=1}^{n-1} \left(\frac{1 - \exp(-\theta d_i)}{\sqrt{1 - \exp(-2\theta d_i)}} \right)^2 \right] = \frac{1}{\sigma_Z^2} \left[1 + \sum_{i=1}^{n-1} \frac{\exp(\theta d_i) - 1}{\exp(\theta d_i) + 1} \right].$$

Proposizione 4.3 *Siano*

$$\mathbf{\Omega} = \mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \frac{\partial \mathbf{L}}{\partial \theta} \mathbf{D}^{\frac{1}{2}} \quad e \quad \mathbf{\Delta} = \mathbf{D}^{-\frac{1}{2}} \frac{\partial \mathbf{D}^{\frac{1}{2}}}{\partial \theta}. \quad (4.13)$$

L'informazione $I_\theta(\mathbf{d}_{\xi_n})$ del parametro della funzione di correlazione θ è data da

$$I_\theta(\mathbf{d}_{\xi_n}) = \sum_{i,j=1}^n \omega_{ij}^2 + 2 \sum_{i,j=1}^n \delta_{ij}^2, \quad (4.14)$$

dove ω_{ij} e δ_{ij} sono l'elemento (i, j) della matrice $\mathbf{\Omega}$ e $\mathbf{\Delta}$, rispettivamente.

Dimostrazione 4.4 *Ricordiamo che*

$$\mathbf{R}^{-1} = \left(\mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \right)^t \left(\mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \right).$$

Pertanto da (4.6) si ha

$$\begin{aligned} I_\theta(\mathbf{d}_{\xi_n}) &= \frac{1}{2} \text{tr} \left[\left(\left(\mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \right)^t \left(\mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \right) \frac{\partial \mathbf{R}}{\partial \theta} \right)^2 \right] \\ &= \frac{1}{2} \text{tr} \left[\left(\left(\mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \right) \frac{\partial \mathbf{R}}{\partial \theta} \left(\mathbf{D}^{-\frac{1}{2}} \mathbf{L}^{-1} \right)^t \right)^2 \right] = \frac{1}{2} \text{tr} \mathbf{B}^2 = \frac{1}{2} \sum_{i,j=1}^n b_{ij}^2 \end{aligned}$$

dove $\mathbf{B} = \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right) \frac{\partial \mathbf{R}}{\partial \theta} \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right)^t$ è una matrice simmetrica dal momento che \mathbf{R} è simmetrica, con generico elemento b_{ij} ($i, j = 1, \dots, n$). Inoltre

$$\begin{aligned} \mathbf{B} &= \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right) \frac{\partial \left[\left(\mathbf{LD}^{\frac{1}{2}}\right) \left(\mathbf{LD}^{\frac{1}{2}}\right)^t \right]}{\partial \theta} \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right)^t \\ &= \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right) \left[\frac{\partial \left(\mathbf{LD}^{\frac{1}{2}}\right)}{\partial \theta} \left(\left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right)^t\right)^{-1} + \left(\mathbf{LD}^{\frac{1}{2}}\right) \frac{\partial \left(\left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right)^t\right)^{-1}}{\partial \theta} \right] \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right)^t \\ &= \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right) \frac{\partial \left(\mathbf{LD}^{\frac{1}{2}}\right)}{\partial \theta} + \left[\left(\mathbf{D}^{-\frac{1}{2}}\mathbf{L}^{-1}\right) \frac{\partial \left(\mathbf{LD}^{\frac{1}{2}}\right)}{\partial \theta} \right]^t = (\mathbf{\Omega} + \mathbf{\Delta}) + (\mathbf{\Omega} + \mathbf{\Delta})^t, \end{aligned}$$

dove $\mathbf{\Delta}$ è una matrice diagonale e $\mathbf{\Omega}$ una matrice triangolare inferiore con elementi sulla diagonale pari a zero. Quindi si ottiene $\mathbf{B} = \mathbf{\Omega} + \mathbf{\Omega}^t + 2\mathbf{\Delta}$, per cui

$$I_\theta(\mathbf{d}_{\xi_n}) = \frac{1}{2} \sum_{i,j=1}^n b_{ij}^2 = \sum_{i,j=1}^n \omega_{ij}^2 + 2 \sum_{i,j=1}^n \delta_{ij}^2.$$

Proposizione 4.5 La somma degli elementi al quadrato della matrice $\mathbf{\Omega} = (\omega_{ij})_{i,j=1,\dots,n}$ definita in (4.13) è data da

$$\sum_{i,j=1}^n \omega_{ij}^2 = \sum_{i=1}^{n-1} \frac{d_i^2}{\exp(2\theta d_i) - 1}. \quad (4.15)$$

Dimostrazione 4.6 La dimostrazione si fa per induzione su n , ovvero la dimensione della matrice quadrata $\mathbf{\Omega}$. La (4.15) è vera per $n = 2$. Infatti, da (4.11) e (4.12) si ha

$$\mathbf{D}_2^{-\frac{1}{2}}\mathbf{L}_2^{-1} = \begin{pmatrix} 1 & 0 \\ \frac{-\exp(-\theta d_1)}{\sqrt{1-\exp(-2\theta d_1)}} & \frac{1}{\sqrt{1-\exp(-2\theta d_1)}} \end{pmatrix}, \quad \frac{\partial \mathbf{L}_2}{\partial \theta} = \begin{pmatrix} 0 & 0 \\ \frac{\partial \exp(-\theta d_1)}{\partial \theta} & 0 \end{pmatrix}$$

$$\mathbf{D}_2^{\frac{1}{2}} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\exp(-2\theta d_1)} \end{pmatrix}$$

per cui

$$\mathbf{\Omega}_2 = \begin{pmatrix} 0 & 0 \\ \frac{-d_1 \exp(-\theta d_1)}{\sqrt{1-\exp(-2\theta d_1)}} & 0 \end{pmatrix} \Rightarrow \sum_{i,j=1}^2 \omega_{ij}^2 = \frac{d_1^2 \exp(-2\theta d_1)}{1-\exp(-2\theta d_1)} = \frac{d_1^2}{\exp(2\theta d_1) - 1}.$$

Le matrici $(n \times n)$ -dimensionali $\mathbf{D}_n^{-\frac{1}{2}}$, $\frac{\partial \mathbf{L}_n}{\partial \theta}$ e $\mathbf{D}_n^{\frac{1}{2}}\mathbf{L}_n^{-1}$ possono essere messe in relazione con le corrispondenti matrici di dimensione $n-1$ semplicemente esprimendo le prime come

matrici a blocchi, ossia

$$\mathbf{D}_n^{\frac{1}{2}} = \left(\begin{array}{c|c} \mathbf{D}_{n-1}^{\frac{1}{2}} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline 0 \ \dots \ 0 & \sqrt{1 - \exp(-2\theta d_{n-1})} \end{array} \right), \quad \frac{\partial \mathbf{L}_n}{\partial \theta} = \left(\begin{array}{c|c} \frac{\partial \mathbf{L}_{n-1}}{\partial \theta} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline \mathbf{m}_{n-1} & 0 \end{array} \right)$$

e

$$\mathbf{D}_n^{-\frac{1}{2}} \mathbf{L}_n^{-1} = \left(\begin{array}{c|c} \mathbf{D}_{n-1}^{-\frac{1}{2}} \mathbf{L}_{n-1}^{-1} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline \mathbf{v}_{n-1} & \frac{1}{\sqrt{1 - \exp(-2\theta d_{n-1})}} \end{array} \right),$$

dove \mathbf{v}_{n-1} and \mathbf{m}_{n-1} sono vettori $(n-1)$ -dimensionali dati da

$$\mathbf{v}_{n-1} = \left(0, \dots, 0, \frac{-\exp(-\theta d_{n-1})}{\sqrt{1 - \exp(-2\theta d_{n-1})}} \right)$$

$$\mathbf{m}_{n-1} = \left(\frac{\partial \exp\left(-\theta \left(\sum_{i=1}^{n-1} d_i\right)\right)}{\partial \theta}, \frac{\partial \exp\left(-\theta \left(\sum_{i=2}^{n-1} d_i\right)\right)}{\partial \theta}, \dots, \frac{\partial \exp(-\theta d_{n-1})}{\partial \theta} \right).$$

Da (4.13) si ottiene

$$\mathbf{\Omega}_n = \left(\begin{array}{c|c} \mathbf{\Omega}_{n-1} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline \mathbf{\Omega}_{n-1}^* & 0 \end{array} \right),$$

dove

$$\mathbf{\Omega}_{n-1}^* = \left[\mathbf{v}_{n-1} \frac{\partial \mathbf{L}_{n-1}}{\partial \theta} + \frac{\mathbf{m}_{n-1}}{\sqrt{1 - \exp(-2\theta d_{n-1})}} \right] \mathbf{D}_{n-1}^{\frac{1}{2}}.$$

Supponendo adesso che la relazione (4.15) sia vera per $n-1$, ovvero

$$\sum_{i,j=1}^{n-1} \omega_{ij}^2 = \sum_{i=1}^{n-2} \frac{d_i^2}{\exp(2\theta d_i) - 1},$$

resta da dimostrare che la somma degli elementi al quadrato del vettore $(n-1)$ -dimensionale

$\mathbf{\Omega}_{n-1}^*$ è pari a

$$\frac{d_{n-1}^2}{\exp(2\theta d_{n-1}) - 1}.$$

Per semplicità di notazione si pone $\phi_i = \exp(-\theta d_i)$, $i = 1, \dots, n-1$ per cui si ha

$$\mathbf{\Omega}_{n-1}^* = \frac{\frac{\partial \phi_{n-1}}{\partial \theta}}{\sqrt{1 - \phi_{n-1}^2}} \left(\prod_{i=1}^{n-2} \phi_i, \sqrt{1 - \phi_1^2} \prod_{i=2}^{n-2} \phi_i, \sqrt{1 - \phi_2^2} \prod_{i=3}^{n-2} \phi_i, \dots, \sqrt{1 - \phi_{n-2}^2} \right)$$

e la somma degli elementi al quadrato di $\mathbf{\Omega}_{n-1}^*$ è pari a

$$\begin{aligned} & \frac{\left(\frac{\partial\phi_{n-1}}{\partial\theta}\right)^2}{1-\phi_{n-1}^2} \left\{ \prod_{i=1}^{n-2} \phi_i^2 + (1-\phi_1^2) \prod_{i=2}^{n-2} \phi_i^2 + (1-\phi_2^2) \prod_{i=3}^{n-2} \phi_i^2 + \dots + (1-\phi_{n-2}^2) \right\} = \\ & = \frac{\left(\frac{\partial\phi_{n-1}}{\partial\theta}\right)^2}{1-\phi_{n-1}^2} = \frac{d_{n-1}^2 \exp(-2\theta d_{n-1})}{1-\exp(-2\theta d_{n-1})} = \frac{d_{n-1}^2}{\exp(2\theta d_{n-1})-1}, \end{aligned}$$

per cui la tesi.

Teorema 4.7 *L'informazione $I_\theta(\mathbf{d}_{\xi_n})$ del parametro della correlazione θ può essere espressa come segue*

$$I_\theta(\mathbf{d}_{\xi_n}) = \sum_{i=1}^{n-1} g(d_i), \quad \text{dove} \quad g(y) = \frac{y^2 (\exp(2\theta y) + 1)}{(\exp(2\theta y) - 1)^2}.$$

Dimostrazione 4.8 *Dalla proposizione 4.3 si ottiene*

$$\mathbf{\Delta} = \mathbf{D}^{-\frac{1}{2}} \frac{\partial(\mathbf{D}^{\frac{1}{2}})}{\partial\theta} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \frac{d_1}{\exp(2\theta d_1)-1} & 0 & \dots & 0 \\ 0 & 0 & \frac{d_2}{\exp(2\theta d_2)-1} & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{d_{n-1}}{\exp(2\theta d_{n-1})-1} \end{pmatrix}.$$

Pertanto

$$\sum_{i,j=1}^n \delta_{ij}^2 = \sum_{i=1}^{n-1} \frac{d_i^2}{(\exp(2\theta d_i) - 1)^2}.$$

Inoltre, per la proposizione 4.5 si ha la tesi. Infatti:

$$I_\theta(\mathbf{d}_{\xi_n}) = \sum_{i=1}^{n-1} \frac{d_i^2}{\exp(2\theta d_i) - 1} + 2 \sum_{i=1}^{n-1} \frac{d_i^2}{(\exp(2\theta d_i) - 1)^2} = \sum_{i=1}^{n-1} \frac{d_i^2 (\exp(2\theta d_i) + 1)}{(\exp(2\theta d_i) - 1)^2}.$$

4.4 I Disegni Ottimi per le Stime Separate di μ e θ

Come anticipato nel paragrafo che descrive la procedura di sviluppo del surrogato, una volta selezionata la tipologia del metamodello, si deve procedere alla sua identificazione, ovvero si devono stimare i parametri a partire da un insieme di valori per le variabili di ingresso (passo 6) scelti in maniera opportuna. Kiefer (1959) descrive la scelta di un piano sperimentale come un problema di decisione in assenza di dati, che richiede prima di tutto la formulazione di una funzione di utilità, o *criterio di ottimalità*, che rifletta lo scopo dell'esperimento e ci permetta di misurare la bontà di un esperimento, o confrontare più

esperimenti tra loro e scegliere il migliore. Definito il criterio di ottimalità, si è in grado di scegliere l'esperimento pianificato nel modo migliore possibile, cioè il disegno ottimo, che è il disegno che massimizza la funzione di utilità. E' necessario dunque stabilire prima di tutto qual'è lo scopo principale dell'esperimento per poi essere in grado di definire un appropriato criterio di ottimalità.

In questo paragrafo ci occupiamo del problema della scelta del piano sperimentale nel caso in cui lo scopo dell'esperimento sia ottenere delle stime "ottime" dei parametri incogniti del metamodello Kriging Ordinario unidimensionale

$$Y(x) = \mu + Z(x), \quad x \in \mathcal{X} \subset \mathbb{R}$$

assumendo una struttura di correlazione esponenziale, ovvero

$$\text{Corr}\{Z(x), Z(x')\} = \exp(-\theta|x - x'|) \quad x, x' \in \mathcal{X}.$$

Dal punto di vista della pianificazione il problema consiste nello scegliere il disegno che minimizza la matrice di varianza covarianza degli stimatori di massima verosimiglianza dei parametri del modello. Ricordiamo che anche in questo contesto essa può essere approssimata attraverso l'inversa della matrice di informazione di Fisher e pertanto l'obiettivo è identificare il disegno che massimizza la FIM. L'ortogonalità delle componenti della matrice di informazione di Fisher (si veda (4.6)) rende lecito lo studio delle stime separate di μ e θ . Inoltre, in termini di disegno ottimo il valore di σ_Z^2 è ininfluenza e pertanto da ora in poi assumeremo $\sigma_Z^2 = 1$ (Zhu e Stein, 2005).

Spesso si è interessati ad ottimizzare l'esperimento rispetto ad opportune funzioni criterio da cui è possibile derivare dei disegni ideali giudicati ottimali dallo sperimentatore. Un esempio in ambito strettamente inferenziale è costituito dai criteri della D- e A-ottimalità, che rappresentano classiche proposte proprie della teoria dei disegni ottimi (Silvey, 1980). Anche se questi criteri sono diffusi nella letteratura legata a osservazioni indipendenti e non correlate, recentemente sono anche stati applicati nel caso di osservazioni correlate (Müller e Pázman (2003), Pázman e Müller (2000), Pepelyshev (2007) e Stehlík (2007)). Tuttavia, noi indugiamo nel proporre un criterio particolare di ottimalità dal momento che i parametri ignoti del nostro modello non sono omogenei (μ caratterizza la struttura della media del processo di Ornstein-Uhlenbeck, mentre θ parametrizza la funzione di correlazione) e l'importanza che viene data ad ognuno di essi dipende dal contesto sperimentale (si veda a riguardo Zimmerman, 2006). Prenderemo in considerazione il criterio della D-ottimalità nel quinto capitolo.

Dai Teoremi 4.1 e 4.7, segue che la matrice di informazione di Fisher è data da

$$\mathbf{I}(\mu, \theta; \mathbf{d}_{\xi_n}) = \begin{pmatrix} 1 + \sum_{i=1}^{n-1} \frac{\exp(\theta d_i) - 1}{\exp(\theta d_i) + 1} & 0 \\ 0 & \sum_{i=1}^{n-1} \frac{d_i^2 (\exp(2\theta d_i) + 1)}{(\exp(2\theta d_i) - 1)^2} \end{pmatrix},$$

e pertanto $\mathbf{I}(\mu, \theta; \mathbf{d}_{\xi_n})$ è invariante per permutazioni delle distanze d_i , per cui ogni disegno composto da n punti che genera medesime distanze dà la stessa informazione.

Teorema 4.9 *Il disegno ottimo $\xi_n^*(\mu)$ per la stima del trend è quello equispaziato, ovvero il disegno composto da punti fra loro equidistanziati.*

Dimostrazione 4.10 *Per la stima del trend dobbiamo scegliere il disegno che massimizza*

$$I_\mu(\mathbf{d}_{\xi_n}) = 1 + \sum_{i=1}^{n-1} f(d_i) = 1 + \sum_{i=1}^{n-1} \frac{\exp(\theta d_i) - 1}{\exp(\theta d_i) + 1}. \quad (4.16)$$

E' facile verificare che $f(\cdot)$ è concava in (d_1, \dots, d_{n-1}) sul simpleso aperto $\sum_{i=1}^{n-1} d_i = 1$. Pertanto dalla proposizione A.2 in Appendice, la funzione

$$I_\mu(\mathbf{d}_{\xi_n}) : (0, 1)^{n-1} \rightarrow \mathbb{R}^+$$

è Schur-concava e il suo massimo è dato da

$$d_i = \frac{1}{n-1}, \quad i = 1, \dots, n-1.$$

Teorema 4.11 *Il disegno ottimo che massimizza l'informazione di Fisher $I_\theta(\mathbf{d}_{\xi_n})$ del parametro della correlazione non esiste all'interno della classe dei disegni ammissibili.*

Dimostrazione 4.12 *Vogliamo scegliere adesso il disegno che massimizza*

$$I_\theta(\mathbf{d}_{\xi_n}) = \sum_{i=1}^{n-1} g(d_i) = \sum_{i=1}^{n-1} \frac{d_i^2 (1 + \exp(2\theta d_i))}{(\exp(2\theta d_i) - 1)^2}. \quad (4.17)$$

Attraverso calcoli assai laboriosi si ottiene che $\frac{\partial^2 g(y)}{\partial y^2}$, la quale è pari a

$$\begin{aligned} \frac{\partial^2 g(y)}{\partial y^2} &= \frac{2}{(\exp(2\theta y) - 1)^4} \{1 + \exp(6\theta y)(2\theta^2 y^2 - 4\theta y + 1) + \\ &+ \exp(2\theta y)(6\theta^2 y^2 + 12\theta y - 1) + \exp(4\theta y)(16\theta^2 y^2 - 8\theta y - 1)\}, \end{aligned} \quad (4.18)$$

risulta non-negativa per ogni $y \in (0; 1)$. Infatti, il fatto che la dodicesima derivata del numeratore di (4.18) è positiva e tutte le sue derivate di ordine inferiore sono non-negative se valutate in $y = 0$, implica che il numeratore in questione è non-negativo. Di conseguenza

la funzione $g(\cdot)$ è convessa in (d_1, \dots, d_{n-1}) sul semplice aperto $\sum_{i=1}^{n-1} d_i = 1$ e $I_\theta(\mathbf{d}_{\xi_n}) : (0, 1)^{n-1} \rightarrow \mathbb{R}^+$ è Schur-convessa, motivo per cui il suo estremo superiore si ottiene nei punti di frontiera. Nel nostro contesto però, $d_i \neq \{0, 1\}$ per $i = 1, \dots, n-1$, e quindi la tesi.

La Schur-convessità di $I_\theta(\mathbf{d}_{\xi_n})$ implica che l'estremo superiore

$$\sup I_\theta(\mathbf{d}_{\xi_n}) = \frac{n-2}{2\theta^2} + \frac{\exp(2\theta) + 1}{(\exp(2\theta) - 1)^2}$$

si ottiene in corrispondenza del disegno con $d_1 = 1$ e $d_2 = d_3 = \dots = d_{n-1} = 0$ (o qualsiasi permutazione delle suddette distanze) per cui il piano sperimentale ideale per la stima di θ , $\tilde{\xi}_n(\theta)$, tende a sovrapporre tutti i punti di osservazione agli estremi dello spazio di disegno, ovvero una distanza tende a uno e le rimanenti a zero. Inoltre, il disegno equispaziato rappresenta la scelta pessima per θ , ossia minimizza $I_\theta(\mathbf{d}_{\xi_n})$ in quanto essa è Schur-convessa.

4.5 Il Disegno della Progressione Geometrica

Esaminando singolarmente la componente di trend μ ed il parametro di correlazione θ , abbiamo dimostrato che il disegno ottimo per la stima del trend è quello equispaziato, mentre per quanto riguarda la stima di θ il piano sperimentale ideale tende a sovrapporre tutti i punti di osservazione agli estremi del dominio. E' stata inoltre dimostrata l'esistenza di un palese contrasto dal punto di vista della precisione di stima dei due parametri, in quanto il disegno ottimo per μ è quello pessimo per θ e viceversa. Alla luce del comportamento conflittuale in questione, l'intento di questo paragrafo consiste in una proposta di una nuova classe di procedure che consentono di ottenere un opportuno compromesso in termini di precisione di stima simultanea di entrambi i parametri, garantendo al tempo stesso alta efficienza singolarmente per ciascuno di essi. In particolare, proponiamo la classe di disegni $\{\xi_n(r)\}_r$ generata dal seguente vettore di distanze

$$\mathbf{d}_{\xi_n(r)} = (k, kr, kr^2, \dots, kr^{n-2}), \quad \text{con } 0 < r \leq 1$$

che chiameremo *Disegno della Progressione Geometrica* (GPD: Geometric Progression Design). In questa classe k è una costante positiva di normalizzazione, a causa del vincolo (4.8), data da

$$k = \begin{cases} \frac{1}{n-1}, & \text{se } r = 1 \\ \frac{1-r}{1-r^{n-1}}, & \text{se } r < 1 \end{cases},$$

ed r è un parametro di tuning che può essere scelto dallo sperimentatore in base all'efficienza desiderata relativa alla stima dei due parametri incogniti del modello.

Si noti che per ogni $n > 2$ le distanze d_i dipendono da n , ovvero

$$d_i \equiv d_{i,n}(r) = \frac{(1-r)r^{i-1}}{1-r^{n-1}} \quad \text{per ogni } i = 1, \dots, n-1. \quad (4.19)$$

Inoltre, se $r = 1$ il disegno della progressione geometrica coincide con il disegno equispaziato, mentre per $r \rightarrow 0$ si ottiene il disegno ideale per θ :

$$\xi_n(1) = \xi_n^*(\mu) \quad \text{e} \quad \lim_{r \rightarrow 0} \xi_n(r) = \tilde{\xi}_n(\theta).$$

In generale si ha:

Teorema 4.13 *Per ogni $n > 2$ fissato e $\theta > 0$*

- (i) $I_\mu(\xi_n(r))$ è crescente rispetto a r ;
- (ii) $I_\theta(\xi_n(r))$ è decrescente rispetto a r .

Dimostrazione 4.14 (i) *Nel caso in cui $r = 1$, l'informazione $I_\mu(\xi_n(r))$ del trend viene massimizzata in corrispondenza del disegno della progressione geometrica per il fatto che quest'ultimo coincide con il disegno equispaziato. Inoltre, per ogni $r \in (0; 1)$ si può dimostrare*

$$\frac{\partial I_\mu(\xi_n(r))}{\partial r} = \sum_{i=1}^{n-1} \frac{\partial f(d_{i,n}(r))}{\partial d_{i,n}(r)} \frac{\partial d_{i,n}(r)}{\partial r} > 0.$$

Infatti, da (4.19), per ogni $n > 2$

$$d_{1,n}(r) > \dots > d_{n-1,n}(r)$$

e

$$0 < \frac{\partial f(d_{1,n}(r))}{\partial d_{1,n}(r)} < \dots < \frac{\partial f(d_{n-1,n}(r))}{\partial d_{n-1,n}(r)},$$

poiché la funzione $f(\cdot)$ è crescente e concava. Per ogni $i = 1 \dots, n-1$

$$\frac{\partial d_{i,n}(r)}{\partial r} = \frac{r^i}{(r-r^n)^2} \{r^{n-1}(n-i) - r^n(n-i-1) + i-1-ri\},$$

dove $\sum_{i=1}^{n-1} [\partial d_{i,n}(r)/\partial r] = 0$ dal vincolo imposto $\sum_{i=1}^{n-1} d_{i,n}(r) = 1$ per ogni $r \in (0; 1]$.

Attraverso semplici passaggi algebrici si dimostra anche

$$\frac{\partial d_{1,n}(r)}{\partial r} < 0, \quad \frac{\partial d_{n-1,n}(r)}{\partial r} > 0$$

e

$$\frac{\partial d_{i,n}(r)}{\partial r} \geq 0 \quad \text{se e solo se} \quad i \geq i^* = \frac{(n-1)r^n - nr^{n-1} + 1}{(1-r)(1-r^{n-1})},$$

dove $1 < i^* < n-1$ dato che

$$1 + (n-1)r^{n-1} < \frac{1-r^n}{1-r} = \sum_{i=0}^{n-1} r^i < n-1 + r^{n-1}.$$

Sia j l'intero più piccolo tale che $\partial d_{i,n}(r)/\partial r \geq 0$ per $i = j, \dots, n-1$. Di conseguenza

$$\begin{aligned} \frac{\partial I_\mu(\xi_n(r))}{\partial r} &= \sum_{i=1}^{j-1} \frac{\partial f(d_{i,n}(r))}{\partial d_{i,n}(r)} \frac{\partial d_{i,n}(r)}{\partial r} + \sum_{i=j}^{n-1} \frac{\partial f(d_{i,n}(r))}{\partial d_{i,n}(r)} \frac{\partial d_{i,n}(r)}{\partial r} \\ &> \frac{\partial f(d_{j,n}(r))}{\partial d_{j,n}(r)} \sum_{i=1}^{n-1} \frac{\partial d_{i,n}(r)}{\partial r} = 0. \end{aligned}$$

per cui la tesi.

(ii) Analogamente, l'informazione $I_\theta(\xi_n(r))$ è decrescente rispetto a r , considerando che la funzione $g(\cdot)$ è decrescente e convessa.

4.5.1 Determinazione della Dimensione Campionaria

Una delle caratteristiche che deve avere un buon piano sperimentale consiste nel ridurre il più possibile il numero dei punti di disegno necessari per acquisire una adeguata quantità di informazione. Per di più, in molti contesti applicativi la dimensione campionaria deve essere necessariamente ridotta, a causa dei costi e dell'elevato dispendio di risorse temporali impiegate dagli esperimenti, come ad esempio nell'ambito dei computer experiments dove le simulazioni sono di solito assai costose e richiedono lunghi tempi di calcolo.

L'intento di questo paragrafo consiste nello studio delle proprietà del disegno della progressione geometrica $\xi_n(r)$ rispetto alla numerosità campionaria n .

Teorema 4.15 Per ogni $r \in (0; 1]$ e $\theta > 0$, l'informazione $I_\mu(\xi_n(r))$ del trend converge per $n \rightarrow \infty$, perciò

$$\lim_{n \rightarrow \infty} n^{-1} I_\mu(\xi_n(r)) = 0.$$

Dimostrazione 4.16 Per $r = 1$, da (4.16) si ottiene

$$I_\mu(\xi_n(1)) = 1 + (n-1) \left(\frac{\exp(\frac{\theta}{n-1}) - 1}{\exp(\frac{\theta}{n-1}) + 1} \right) \rightarrow 1 + \frac{\theta}{2} \quad \text{per } n \rightarrow \infty.$$

Per $0 < r < 1$ e n che tende all'infinito

$$d_{i,n}(r) \rightarrow \tilde{d}_i(r) = (1-r)r^{i-1} \quad \text{per ogni } i$$

e quindi

$$\lim_{n \rightarrow \infty} f(d_{i,n}(r)) = f(\tilde{d}_i(r)) = \frac{\exp(\theta(1-r)r^{i-1}) - 1}{\exp(\theta(1-r)r^{i-1}) + 1}.$$

In aggiunta, per ogni $n > 2$

$$0 < f(d_{i,n}(r)) \leq f(d_{i,2}(r)) = f(r^{i-1})$$

dove $\sum_{i=1}^{\infty} f(r^{i-1}) < \infty$ giacché la funzione $f(\cdot)$ è strettamente crescente e

$$\lim_{i \rightarrow \infty} \frac{f(r^i)}{f(r^{i-1})} = r < 1.$$

Di conseguenza, dal Test M di Weierstrass (Billingsley, 1995)

$$\lim_{n \rightarrow \infty} I_{\mu}(\xi_n(r)) = 1 + \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} f(d_{i,n}(r)) = 1 + \sum_{i=1}^{\infty} f(\tilde{d}_i(r)). \quad (4.20)$$

Dato che la sequenza $\{\tilde{d}_i(r)\}_{i \in \mathbb{N}}$ è strettamente decrescente si ha $f(\tilde{d}_i(r)) > f(\tilde{d}_{i+1}(r)) > 0$ per ogni i e

$$\lim_{i \rightarrow \infty} \frac{f(\tilde{d}_{i+1}(r))}{f(\tilde{d}_i(r))} = r < 1;$$

e pertanto $\sum_{i=1}^{\infty} f(\tilde{d}_i(r))$ converge e (4.20) implica $\lim_{n \rightarrow \infty} I_{\mu}(\xi_n(r)) < \infty$.

Questo effetto paradossale può essere spiegato dal fatto che le osservazioni sono correlate: la correlazione riduce la necessità di prendere ulteriori osservazioni e questo comportamento è conosciuto nella letteratura relativa ai dati correlati come *effetto di saturazione* (si veda ad esempio Anisimov et al, 2007).

La Figura (4.1) fornisce i grafici di $I_{\mu}(\xi_n(r))$ in funzione di n per vari valori di r ($r = 0.3$ (rosso), $r = 0.5$ (verde), $r = 0.7$ (blu)) e θ ($\theta = 0.5$ (sinistra), $\theta = 1$ (centrale), $\theta = 5$ (destra)). E' evidente che l'informazione del parametro di trend cresce nel caso di dimensione campionaria contenuta, mentre superato un certo valore critico, il cosiddetto *cut-off*, l'informazione è costante.

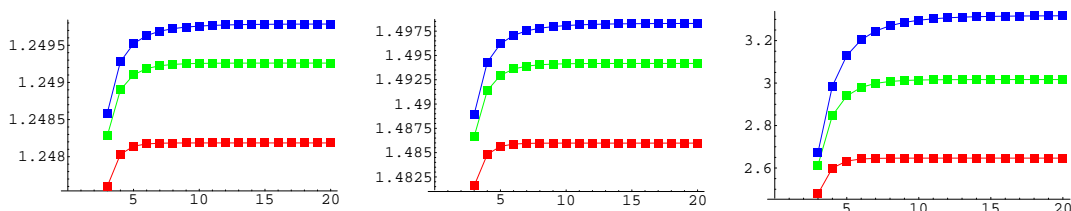


Figura 4.1: $I_{\mu}(\xi_n(r))$ in funzione di n .

Contrariamente al comportamento di $I_\mu(\xi_n(r))$, l'informazione legata al parametro di correlazione θ cresce al crescere della dimensione campionaria come mostra il seguente risultato:

Teorema 4.17 Per ogni $r \in (0; 1]$ e $\theta > 0$, al tendere di n all'infinito

$$(i) \quad I_\theta(\xi_n(r)) \rightarrow \infty,$$

$$(ii) \quad \frac{I_\theta(\xi_n(r))}{n} \rightarrow \frac{1}{2\theta^2}.$$

Dimostrazione 4.18 (i) Per $r = 1$, (4.17) implica

$$I_\theta(\xi_n(1)) = \frac{\left(1 + \exp\left(\frac{2\theta}{n-1}\right)\right)}{(n-1) \left(\exp\left(\frac{2\theta}{n-1}\right) - 1\right)^2}$$

e pertanto per $n \rightarrow \infty$

$$I_\theta(\xi_n(1)) \rightarrow \infty \quad e \quad \frac{I_\theta(\xi_n(1))}{n} \rightarrow \frac{1}{(2\theta^2)}.$$

Per ogni $r \in (0; 1)$, la dimostrazione di (i) è analoga a quella relativa a $I_\mu(\xi_n(r))$ (si veda a riguardo la dimostrazione del Teorema 4.15): per n che tende all'infinito l'informazione di Fisher $I_\theta(\xi_n(r))$ diverge, dato che $g(\cdot)$ è una funzione decrescente.

(ii) Innanzitutto si noti che per ogni $n > 2$

$$0 \leq g(r^{i-1}) \leq g(d_{i,n}(r)) \leq g((1-r)r^{i-1}) \quad \forall i.$$

Di conseguenza,

$$\frac{1}{n} \sum_{i=1}^{n-1} g(r^{i-1}) \leq \frac{1}{n} \sum_{i=1}^{n-1} g(d_{i,n}(r)) \leq \frac{1}{n} \sum_{i=1}^{n-1} g((1-r)r^{i-1}).$$

Considerato che

$$g(r^{i-1}) = \left(\frac{r^{i-1}}{\exp(2\theta r^{i-1}) - 1}\right)^2 (1 + \exp(2\theta r^{i-1})) \rightarrow \frac{1}{2\theta^2} \quad \text{per } i \rightarrow \infty$$

e

$$g((1-r)r^{i-1}) = \left(\frac{(1-r)r^{i-1}}{\exp(2\theta(1-r)r^{i-1}) - 1}\right)^2 (1 + \exp(2\theta(1-r)r^{i-1}))$$

che tende a $\frac{1}{2\theta^2}$ per $i \rightarrow \infty$, dal lemma di Cesàro (Cesàro, 1890) si ha

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n-1} g(r^{i-1}) = \frac{1}{2\theta^2} \quad e \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n-1} g((1-r)r^{i-1}) = \frac{1}{2\theta^2},$$

per cui

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n-1} g(d_{i,n}(r)) = \frac{1}{2\theta^2}.$$

La Figura (4.2) evidenzia che l'aumento della dimensione campionaria ha un impatto molto importante per θ specialmente nel caso in cui la correlazione assume valori elevati, ossia per valori piccoli di θ .

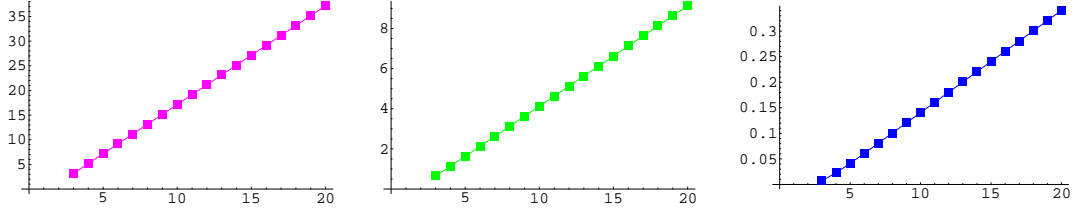


Figura 4.2: $I_\theta(\xi_n(0.3))$ in funzione di n per $\theta = 0.5$ (sinistra), $\theta = 1$ (centrale) e $\theta = 5$ (destra).

4.5.2 Efficienza del Disegno della Progressione Geometrica

Una misura di efficienza del disegno della progressione geometrica $\xi_n(r)$ per la stima di ciascuno dei parametri incogniti del modello sotto esame viene data dal rapporto tra l'informazione e il suo massimo (o estremo superiore nel caso di θ), ovvero

$$\text{Eff}_\mu(\xi_n(r)) = \frac{I_\mu(\xi_n(r))}{I_\mu(\xi_n^*(\mu))} \quad \text{e} \quad \text{Eff}_\theta(\xi_n(r)) = \frac{I_\theta(\xi_n(r))}{I_\theta(\tilde{\xi}_n(\theta))}.$$

Tale misura ci permette di studiare il comportamento del Disegno della Progressione Geometrica in funzione di r e θ .

Nella Figura (4.3) sono rappresentati i grafici delle suddette efficienze in funzione di $r \in (0; 1]$: per sottolineare le performance del disegno della progressione geometrica abbiamo selezionato tre valori diversi per θ e abbiamo fissato $n = 15$ considerato che il cut-off inerente a μ è prossimo a questo valore per ogni r e θ . Questa figura evidenzia il fatto che al decrescere di r , $\text{Eff}_\mu(\xi_n(r))$ decresce, mentre $\text{Eff}_\theta(\xi_n(r))$ cresce. Inoltre, per valori piccoli di θ , ovvero per osservazioni fortemente correlate, il disegno della progressione geometrica è estremamente efficiente, garantendo ottime performance per entrambi i parametri, mentre al crescere di θ le efficienze del GPD decrescono (si vedano anche le Figure (4.4) e (4.5)).

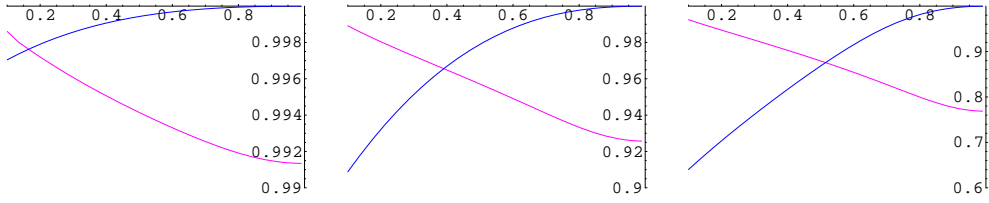


Figura 4.3: Grafici dell'efficienza riguardante μ (blu) e θ (rosa) del GPD $\xi_{15}(r)$ con $\theta = 0.5$ (sinistra), $\theta = 2$ (centrale) e $\theta = 5$ (destra).

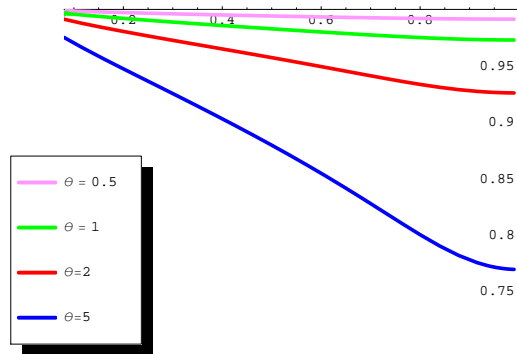


Figura 4.4: Grafici dell'efficienza riguardante θ del GPD $\xi_{15}(r)$ al variare di r .

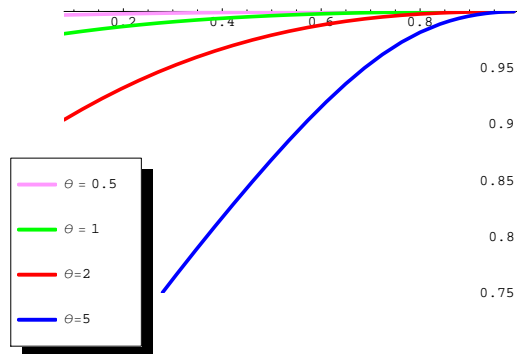


Figura 4.5: Grafici dell'efficienza riguardante μ del GPD $\xi_{15}(r)$ al variare di r .

Capitolo 5

Disegni Sequenziali per il Modello di Decadimento Esponenziale

In questo capitolo prendiamo in esame un diverso modello parzialmente lineare, ossia il cosiddetto Modello di Decadimento Esponenziale (MDE), che trova le sue applicazioni principalmente nel campo della farmacocinetica. Vengono formulate due proposte di esperimento per la stima dei suoi parametri attraverso delle procedure sequenziali. La prima si basa sulla massima verosimiglianza; la seconda sfrutta la caratteristica della parziale linearità. Le proprietà di queste proposte vengono studiate per mezzo di una serie di simulazioni.

5.1 Il Modello di Decadimento Esponenziale

Il modello, oggetto di studio di questo capitolo, è il modello di decadimento esponenziale dato da

$$E(Y|x) = \tau(\boldsymbol{\beta}, x) = \mu \exp(-\theta x), \quad V(Y|x) = \sigma^2, \quad (5.1)$$

con $\boldsymbol{\beta} = (\mu, \theta)^t$, σ^2 costante, Y distribuito secondo una normale, $\mu \geq 0$, $\theta > 0$ e $x \in \mathcal{X} \subset \mathbb{R}^+$. Ovviamente il modello (5.1) appartiene alla classe dei modelli parzialmente lineari in quanto il valore atteso della risposta Y è lineare rispetto a μ . Tale modello viene ampiamente utilizzato nel campo della farmacocinetica la quale ha come obiettivo lo studio dell'evoluzione temporale della concentrazione di un farmaco nell'organismo (Landaw e DiStefano, 1984). Inoltre, sotto condizioni ideali, questo modello descrive il decadimento

del virus dell'epatite B (hepatitis B virus: HBV) e C (hepatitis C virus: HCV) durante il trattamento al variare del tempo x (Nowak et al, 1996, Neumann et al, 2000).

Essendo il modello di decadimento esponenziale un modello di regressione non lineare con σ^2 costante e Y distribuito normalmente, ed essendo il valore atteso della risposta derivabile rispetto sia a μ che a θ , una soluzione per la stima dei parametri del modello è quella suggerita da Box e Lucas (1959), ovvero rendere il problema del disegno equivalente a quello per un modello lineare approssimando la funzione $\tau(\boldsymbol{\beta}, x) = \mu \exp(-\theta x)$ con una funzione lineare sviluppandola in serie di Taylor al primo ordine intorno a una stima iniziale dei parametri in modo da poter applicare la teoria vista per i modelli lineari. In particolare, si ottiene che il vettore delle derivate parziali è dato da

$$\mathbf{f}(\boldsymbol{\beta}, x) = \nabla \tau(\boldsymbol{\beta}, x) = \begin{pmatrix} \frac{\partial[\mu \exp(-\theta x)]}{\partial \mu} \\ \frac{\partial[\mu \exp(-\theta x)]}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \exp(-\theta x) \\ -\mu x \exp(-\theta x) \end{pmatrix}$$

e se indichiamo con $\boldsymbol{\beta}_0 = (\mu_0, \theta_0)^t$ la stima iniziale, l'espansione di Taylor al primo ordine assume la seguente forma:

$$E(Y) = \tau(\boldsymbol{\beta}, x) = \tau(\boldsymbol{\beta}_0, x) + \nabla \tau(\boldsymbol{\beta}_0, x)^t (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \dots \simeq \tau(\boldsymbol{\beta}_0, x) + \mathbf{f}(\boldsymbol{\beta}_0, x)^t (\boldsymbol{\beta} - \boldsymbol{\beta}_0)$$

ovvero

$$E(Y) - \tau(\boldsymbol{\beta}_0, x) \simeq \mathbf{f}(\boldsymbol{\beta}_0, x)^t (\boldsymbol{\beta} - \boldsymbol{\beta}_0) = (\mu - \mu_0) \exp(-\theta_0 x) + (\theta - \theta_0) (-\mu_0 x \exp(-\theta_0 x))$$

per cui si può applicare la metodologia legata ai modelli lineari, tenendo conto che la suddetta approssimazione dipende dalla stima iniziale.

Come abbiamo già accennato, il criterio più comunemente usato nel caso in cui siamo interessati alla stima dei parametri di un modello è la D-ottimalità. Adottando questo criterio, è facile vedere che il disegno localmente D-ottimo per il modello sotto esame non dipende da μ . Infatti, si noti che

$$\nabla \tau(\boldsymbol{\beta}, x) = \begin{pmatrix} \exp(-\theta x) \\ -\mu x \exp(-\theta x) \end{pmatrix} = \mathbf{A}(\mu) g(\theta, x)$$

dove

$$\mathbf{A}(\mu) = \begin{pmatrix} 1 & 0 \\ 0 & \mu \end{pmatrix} \text{ e } g(\theta, x) = \begin{pmatrix} \exp(-\theta x) \\ -x \exp(-\theta x) \end{pmatrix}$$

e pertanto la matrice di informazione per il disegno φ può essere espressa come segue:

$$\mathbf{M}(\varphi) = \int_{\mathcal{X}} \mathbf{A}(\mu)g(\theta, x)g(\theta, x)^t \mathbf{A}(\mu)^t \varphi(dx) = \mathbf{A}(\mu) \left\{ \int_{\mathcal{X}} g(\theta, x)g(\theta, x)^t \varphi(dx) \right\} \mathbf{A}(\mu)^t$$

per cui, passando al determinante, si ha

$$\det[\mathbf{M}(\varphi)] = \det[\mathbf{A}(\mu)]^2 \det \left\{ \int_{\mathcal{X}} g(\theta, x)g(\theta, x)^t \varphi(dx) \right\}$$

e quindi la scelta del disegno che massimizza tale determinante non dipende da μ .

Box e Lucas (1959) hanno fornito il disegno localmente D-ottimo per la stima dei due parametri presenti nel modello (5.1) nel caso in cui $\mathcal{X} = [x_{\min}, x_{\max}]$ con $x_{\max} \geq x_{\min} + \frac{1}{\theta}$, mentre Walter e Pronzato (1997) per $x \in [0, \infty)$. Han e Chaloner (2003) hanno unificato e generalizzato i risultati ottenuti dai suddetti autori attraverso un teorema che include anche il caso di domini con $x_{\max} < x_{\min} + \frac{1}{\theta}$:

Teorema 5.1 *Il disegno localmente D-ottimo per il modello di decadimento esponenziale (5.1) è dato da due punti di supporto, x_{\min} e $\min\{x_{\max}, x_{\min} + \frac{1}{\theta}\}$, di ugual peso.*

Per quanto riguarda l'approccio Bayesiano, Dette e Neugebauer (1997), e Walter e Pronzato (1997) hanno preso in considerazione disegni D-ottimi Bayesiani con due punti di supporto.

5.2 Procedure Sequenziali per il Modello di Decadimento Esponenziale

La sperimentazione sequenziale rappresenta sicuramente una tra le più recenti ed innovative proposte nell'ambito della teoria del disegno degli esperimenti. Essa, come abbiamo sottolineato più volte, assume un'importanza fondamentale in ambito non lineare in quanto rappresenta una soluzione naturale ai problemi di ottimalità locale. La pianificazione di un esperimento sequenziale, ossia la specificazione del processo di acquisizione delle osservazioni, viene scelta a priori dallo sperimentatore e comprende l'adozione del piano di campionamento delle unità statistiche, la scelta di una regola di determinazione dei punti di disegno e l'utilizzo di una eventuale regola di arresto. In generale, la scelta del disegno sperimentale da adottare dipende da molteplici fattori, tra cui gli obiettivi stessi legati all'investigazione, il modello statistico adottato a descrizione del fenomeno di interesse e l'approccio inferenziale che si è deciso di utilizzare.

Occorre tuttavia osservare che nel caso di esperimenti sequenziali in cui le decisioni sperimentali dipendono di volta in volta dalle risposte osservate, è richiesto un approccio critico all'analisi dei dati ottenuti, sia nel caso di piccoli campioni che nel caso asintotico, dato che i metodi inferenziali classici possono non essere validi.

5.2.1 Procedura basata sulla Massima Verosimiglianza

Nell'ambito degli esperimenti sequenziali per la stima dei parametri presenti in un modello non lineare, Chaudhuri e Mykland (1993) hanno suggerito una particolare metodologia che abbiamo illustrato nel primo capitolo. Gli stimatori che si ottengono attraverso la loro proposta sono consistenti ed asintoticamente normali sotto certe condizioni che, però, sono di difficile verificabilità. Pertanto, risulta conveniente trovare degli esperimenti particolari caso per caso.

Per quanto riguarda il modello di decadimento esponenziale, la prima soluzione che proponiamo è composta da due fasi e in entrambe il metodo di stima utilizzato è la massima verosimiglianza. Inoltre, la seconda fase consiste in un piano sperimentale a più stadi che utilizza il disegno localmente D-ottimo, cioè il piano sperimentale all' i -esimo stadio è il disegno localmente D-ottimo in cui il valore del parametro θ viene sostituito dalla stima ottenuta nel piano precedente. La soluzione in questione può essere descritta attraverso i seguenti passaggi:

Sia n il numero totale dei punti di disegno e φ il disegno localmente D-ottimo fornito dal Teorema 5.1.

FASE PRELIMINARE:

Vengono scelti $m < n$ punti nella maniera che lo sperimentatore ritiene più opportuna e sulla base delle osservazioni ottenute vengono stimati i parametri μ e θ attraverso il metodo della massima verosimiglianza.

FASE SEQUENZIALE:

1. *Esperimento*: per quanto riguarda la scelta dei punti di osservazione successivi, si considera il disegno localmente D -ottimo, φ , per la stima simultanea dei parametri presenti nel modello. Tale disegno dipende solamente da θ e lo si valuta nella stima corrente del parametro in questione;
2. *aggiornamento delle stime*: sulla base delle osservazioni ottenute precedentemente

(incluse quelle fornite dalla fase preliminare) vengono stimati i parametri attraverso il metodo della massima verosimiglianza e si ritorna al passo 1 avendo aggiornato le stime.

Abbiamo implementato con Matlab la procedura in questione per il modello di decadimento esponenziale supponendo di restringere lo spazio di disegno \mathcal{X} all'intervallo $(0; M]$ con M molto grande e impostando i seguenti vincoli:

- abbiamo posto $\sigma^2 = 1$;
- abbiamo considerato $n = 100$ e $m = 10$;
- gli m punti della prima fase della procedura sono stati distribuiti nell'intervallo $(0; 1]$;
- per poter valutare la variabile di risposta Y e calcolare l'errore quadratico medio sono stati adottati diversi valori come valori veri dei parametri.

L'algoritmo è stato ripetuto per scenari diversi variando i valori veri dei parametri presenti nel modello e tenendo fisso il numero di iterazioni¹ riguardanti la fase sequenziale della metodologia, che abbiamo assunto pari a 45 dato che il disegno localmente D-ottimo è composto da due punti di supporto e $n - m = 90$. I risultati ottenuti saranno presentati nel paragrafo 5.3 (si veda la Tabella 5.1).

5.2.2 Procedura basata sulla Parziale Linearità

Per la stima dei parametri μ e θ del modello di decadimento esponenziale (5.1) si può pensare di sfruttare la proprietà dei modelli parzialmente lineari, cioè la linearità del valore atteso della risposta rispetto a μ . Partendo da questa considerazione, proponiamo la seguente procedura iterativa:

1. Viene adottato inizialmente un disegno con m punti di supporto;
2. si valuta la variabile di risposta Y nei punti di disegno ottenuti al passo 1;
3. sulla base delle osservazioni ottenute precedentemente vengono stimati sia μ che θ attraverso il metodo della massima verosimiglianza. Le stime in questione saranno indicate con (μ_0, θ_0) ;

¹Con il termine *iterazioni* intendiamo il numero di volte che si ripetono i passi delle procedure.

4. la scelta dei punti di osservazione successivi viene data dal disegno localmente D -ottimo (si veda il Teorema 5.1) per la stima simultanea di entrambi i parametri μ e θ . Tale disegno dipende solo da θ e lo si valuta nella stima corrente di θ , ossia in θ_0 ;
5. si valuta la variabile di risposta Y nei punti di disegno ottenuti al passo 4;
6. sulla base di θ_0 e delle osservazioni ottenute precedentemente viene stimato μ attraverso il metodo dei minimi quadrati;
7. sulla base delle osservazioni ottenute precedentemente e la stima corrente di μ viene stimato θ attraverso il metodo della verosimiglianza profilo (Severini, 2000) e si ritorna al passo 4 avendo aggiornato θ_0 .

Osservazione 5.2 *Si noti che nella suddetta procedura le stime di μ via via ottenute vengono utilizzate ogni volta al punto 7.*

Abbiamo implementato la procedura di cui sopra utilizzando Matlab e impostando i vincoli precedentemente elencati.

5.3 Confronti

In questo paragrafo si vogliono fare dei confronti tra la prima procedura proposta e quella che sfrutta la parziale linearità del modello preso in considerazione. Entrambi gli algoritmi verranno ripetuti per scenari diversi variando i valori veri di μ e θ e tenendo fisso il numero di iterazioni che come abbiamo già accennato è pari a 45, al fine di valutare attraverso l'errore quadratico medio le performance di ciascuna procedura ed effettuare i confronti tra le due. Vogliamo sottolineare di nuovo che nel caso di esperimenti sequenziali in cui le decisioni sperimentali dipendono di volta in volta dalle risposte osservate i metodi inferenziali classici possono non essere validi dal momento che non si hanno osservazioni indipendenti. Le Tabelle 5.1, 5.2 e 5.3 contengono i risultati ottenuti dopo 500 replicazioni.

Come si evince dalla Tabella 5.3, la procedura basata sulla parziale linearità presenta un errore quadratico medio legato al parametro μ fortemente minore rispetto a quella che si basa sulla massima verosimiglianza, mentre si nota un leggero miglioramento per quanto riguarda θ . In aggiunta, la seconda metodologia proposta in questo capitolo riduce la dimensionalità del problema di ottimizzazione inerente al metodo di stima dato che si passa da una massimizzazione in due dimensioni a una che coinvolge una sola variabile.

Tabella 5.1: Procedura sequenziale basata sulla massima verosimiglianza

Valori veri (μ, θ)	(5,1)	(11,0.25)	(50,0.1)	(70,9)	(80,3)
$MSE_{ML}(\mu)$	0.0186	0.018	0.0187	0.0192	0.0189
$MSE_{ML}(\theta)$	0.0068	$8.3491 \cdot 10^{-5}$	$1.3143 \cdot 10^{-6}$	0.0028	$2.2859 \cdot 10^{-4}$

Tabella 5.2: Procedura sequenziale basata sulla parziale linearità

Valori veri (μ, θ)	(5,1)	(11,0.25)	(50,0.1)	(70,9)	(80,3)
$MSE_{PL}(\mu)$	$4.4574 \cdot 10^{-4}$	$9.5658 \cdot 10^{-4}$	$3.4818 \cdot 10^{-4}$	$3.0407 \cdot 10^{-4}$	$4.3081 \cdot 10^{-4}$
$MSE_{PL}(\theta)$	0.0049	$7.7 \cdot 10^{-5}$	$6.1227 \cdot 10^{-7}$	0.0021	$2.0305 \cdot 10^{-4}$

Tabella 5.3: Rapporto tra l'errore quadratico medio ricavato dalla prima (MSE_{ML}) e dalla seconda procedura (MSE_{PL}) per μ e θ

Valori veri (μ, θ)	(5,1)	(11,0.25)	(50,0.1)	(70,9)	(80,3)
$\frac{MSE_{ML}(\mu)}{MSE_{PL}(\mu)}$	41.7284	18.8170	53.7079	63.1434	43.8708
$\frac{MSE_{ML}(\theta)}{MSE_{PL}(\theta)}$	1.3878	1.0843	2.1466	1.3333	1.1258

Questo tipo di guadagno viene confermato anche dal tempo computazionale necessario per ciascun algoritmo: la procedura basata sulla parziale linearità richiede per la sua implementazione un decimo del tempo richiesto dal primo piano sperimentale presentato.

5.4 Conclusioni

In questo capitolo abbiamo proposto due soluzioni sequenziali e abbiamo effettuato dei confronti attraverso i quali si evince che la procedura basata sulla caratteristica della linearità del valore atteso della risposta rispetto ad un parametro presente nel modello risulta migliore sotto vari punti di vista. Siccome le proprietà del piano sperimentale in questione sono state studiate per mezzo di una serie di simulazioni, uno dei possibili sviluppi futuri di questo lavoro è quello di analizzare matematicamente tali proprietà. In più, sarebbe interessante generalizzare tale metodologia per un modello di regressione parzialmente lineare qualsiasi, individuando come prima cosa le condizioni sotto le quali gli stimatori ottenuti sono consistenti. Riteniamo che la proposta in questione può avere delle ottime performance nel caso di modelli di questo tipo in cui il valore atteso della risposta risulta

- lineare rispetto a un parametro vettoriale di dimensione elevata e
- non lineare rispetto a un parametro vettoriale di dimensione contenuta.

Appendice A

La Maggiorizzazione

Come descritto nel testo di Marshall e Olkin (1979), dati due vettori $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$, si dice che \mathbf{x} è maggiorizzato da \mathbf{y} (indicandolo $\mathbf{x} \prec \mathbf{y}$) se:

$$\begin{cases} \sum_{i=1}^k x_{[i]} \leq \sum_{i=1}^k y_{[i]}, & k = 1, \dots, n-1 \\ \sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]} \end{cases}$$

dove con $x_{[i]}$ si denota la componente i -esima del vettore ottenuto ordinando in senso non crescente le componenti del vettore \mathbf{x} , cioè

$$x_{[1]} \geq \dots \geq x_{[n]}.$$

Un semplice esempio di maggiorizzazione è il seguente

$$\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \prec (a_1, \dots, a_n) \prec (1, 0, \dots, 0)$$

per ogni $a_i \geq 0$, $\sum_{i=1}^n a_i = 1$.

Definizione A.1 Una funzione ϕ a valori reali definita in un insieme $\mathbb{A} \subset \mathbb{R}^n$ viene definita *Schur-convessa* in \mathbb{A} se

$$\mathbf{x} \prec \mathbf{y} \quad \text{in } \mathbb{A} \Rightarrow \phi(\mathbf{x}) \leq \phi(\mathbf{y})$$

Analogamente, ϕ viene definita *Schur-concava* in \mathbb{A} se

$$\mathbf{x} \prec \mathbf{y} \quad \text{in } \mathbb{A} \Rightarrow \phi(\mathbf{x}) \geq \phi(\mathbf{y})$$

Proposizione A.2 Sia $\mathbb{I} \subset \mathbb{R}$ un intervallo e $g : \mathbb{I} \rightarrow \mathbb{R}$ una funzione convessa. Allora si ha che la funzione

$$\phi(\mathbf{x}) = \sum_{i=1}^n g(x_i)$$

è Schur-convessa in \mathbb{I}^n . Di conseguenza, $\mathbf{x} \prec \mathbf{y}$ in $\mathbb{I}^n \Rightarrow \phi(\mathbf{x}) \leq \phi(\mathbf{y})$.

Bibliografia

- [1] Abt M., Welch W.J. (1998). *Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian stochastic processes*, The Canadian Journal of Statistics, **26**: 127-137.
- [2] An J., Owen A.B. (2001). *Quasi-regression*, Journal of Complexity, **17**: 588-607.
- [3] Anisimov V.V., Fedorov V.V., Leonov S.L. (2007). *Optimal Design of Pharmacokinetic Studies Described by Stochastic Differential Equations*, López-Fidalgo J., Rodríguez-Díaz J.M. and Torsney B. Eds, mODa 8-Advances in Model-Oriented Design and Analysis, Contributions to Statistics, 9-16, Physica-Verlag.
- [4] Antoniadis A., Fan J. (2001). *Regularization of wavelets approximations (with discussion)*, Journal of the American Statistical Association, **96**: 939-967.
- [5] Atkinson A.C., Donev A.N. (1992). *Optimum Experimental Designs*, Clarendon Press, Oxford.
- [6] Bates R. A., Buck R.J., Riccomagno E., Wynn H.P. (1996). *Experimental design and observation for large systems (with discussion)*, Journal of the Royal Statistical Society (B), **58**: 77-94.
- [7] Bates R. A., Giglio B., Wynn H.P. (2003). *A global selection procedure for polynomial interpolators*, Technometrics, **45**: 246-255.
- [8] Billingsley P. (1995). *Probability and Measure* (3rd ed.), Wiley & Sons.
- [9] Blanning R. W. (1975). *The construction and implementation of metamodels*, Simulation, **24**: 177-184.

- [10] Bogaert P., Russo D. (1999). *Optimal spatial sampling design for the estimation of the variogram based on a least squares approach*, Water Resources Research, **35**: 1275-1289.
- [11] Box G.E.P., Lucas H.L. (1959). *Designs of Experiments in Non-linear Situations*, Biometrika, **49**: 77-90.
- [12] Cesàro E. (1890). *Sur la multiplication des séries*, Bulletin des Sciences Mathématiques, s. II, **14**: 114-120.
- [13] Chaudhuri P., Mykland P.A. (1993). *Nonlinear Experiments: Optimal Design and Inference Based on Likelihood*, Journal of the American Statistical Association **88**: 538-546.
- [14] Cressie N.A.C., Hawkins D.M. (1980). *Robust estimation of the variogram: I*, Mathematical Geology, **12**: 115-125.
- [15] Dette H., Neugebauer H.M. (1997). *Bayesian D-optimal designs for exponential regression models*, Journal of Statistical Planning and Inference, **60**: 331-349.
- [16] Fang K.T., Hickernell F.J. (1995). *The uniform design and its applications*, Bulletin of The International Statistical Institute, 339-349, Beijing.
- [17] Fang K.T., Lin D.K., Winker P., Zhang Y. (2000). *Uniform Design: Theory and Application*, Technometrics, **42**: 237-248.
- [18] Fang K.T., Runze L., Sudjianto A. (2006). *Design and Modeling for Computer Experiments*, Chapman & Hall, UK.
- [19] Fisher R.A. (1935). *The Design of Experiments*, Edinburgh: Oliver & Boyd.
- [20] Ford I., Torsney B., Wu C.F.J. (1992). *The Use of a Canonical Form in the Construction of Locally Optimal Designs for Non-Linear Problems*, Journal of the Royal Statistical Society (B), **54**: 569-583.
- [21] Friedman J.H. (1991). *Multivariate adaptive regression splines*, The Annals of Statistics, **19**: 1-141.
- [22] Gupta A., Ding Y., Xu L. (2006). *Optimal Parameter Selection for Electronic Packaging Using Sequential Computer Simulations*, Journal of Manufacturing Science and Engineering, **128**: 705-715.

- [23] Han C., Chaloner K. (2003). *D- and c-optimal designs for exponential regression models used in viral dynamics and other applications*, Journal of Statistical Planning and Inference, **115**: 585-601.
- [24] Harville D.A. (1997). *Matrix Algebra From a Statistician's Perspective*, Springer-Verlag, New York.
- [25] Hill P.D.H. (1980). *D-Optimal Designs For Partially Nonlinear Regression Models*, Technometrics, **22**: 275-276.
- [26] Iman R.L., Conover W.J. (1982). *A distribution-free approach to inducing rank correlation among input variables*, Communications in Statistics, Series B, **11**: 311-334.
- [27] Johnson M.E., Moore L.M., Ylvisaker D. (1990). *Minimax and maxmin distance design*, Journal of Statistical Planning and Inference, **26**: 131-148.
- [28] Kiefer J. (1959). *Optimum Experimental Designs*, Journal of the Royal Statistical Society (B), **21**: 272-319.
- [29] Kiefer J., Wolfowitz J. (1960). *The equivalence of two extremum problems*, Canadian Journal of Mathematics, **12**: 363-366.
- [30] Kleijnen J.P.C., Sargent R.G. (2000). *A methodology for fitting and validating metamodels in simulation*, European Journal of Operational Research, **120**: 14-29.
- [31] Kleijnen J.P.C., van Beers W.C.M. (2004). *Application-driven sequential designs for simulation experiments: Kriging metamodelling*, Journal of the Operational Research Society, **55**: 876-883.
- [32] Landaw E.W., DiStefano, J.J. (1984). *Multieponential, multicompartmental, and non-compartmental modelling II. Data analysis and statistical considerations*, The American Journal of Physiology, **246**: R665-R677.
- [33] Li R. (2002). *Model selection for analysis of uniform design and computer experiment*, International Journal of Reliability, Quality and Safety Engineering, **9**: 305-315.
- [34] Li R., Sudjianto A. (2005). *Analysis of computer experiments using penalized likelihood in Gaussian Kriging models*, Technometrics, **47**: 111-120.

- [35] Marshall A.W., Olkin I. (1979). *Inequalities: Theory of Majorization and Its Applications*, Academic Press.
- [36] Matheron G. (1962). *Traité de Géostatistique appliquée, Vol. 1, Mémoires du Bureau de Recherches Géologiques et Minières*, Editions Bureau de Recherches Géologiques et Minières, vol. 24. Paris.
- [37] McCullagh P., Nelder J.A. (1989). *Generalized Linear Models*, Chapman & Hall (second edition).
- [38] McKay M.D., Beckman R.J., Conover W.J. (1979). *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics, **21**: 239-245.
- [39] Morris M.D., Mitchell T.J. (1995). *Exploratory designs for computational experiments*, Journal of Statistical Planning and Inference, **43**: 381-402.
- [40] Müller W.G., Pázman A. (2003). *Measures for designs in experiments with correlated errors*, Biometrika, **90**: 423-434.
- [41] Müller W.G., Zimmerman D.L. (1999). *Optimal designs for variogram estimation*, Environmetrics, **10**: 23-37.
- [42] Neumann A.U., Lam N.P., Dahari H., Davidian M., Wiley T.E., Mika B.P., Perelson A.S., Layden T.J. (2000). *Differences in Viral Dynamics between Genotypes 1 and 2 of Hepatitis C Virus*, The Journal of Infectious Diseases, **182**: 28-35.
- [43] Nowak M.A., Bonhoeffer S., Hill M.A., Boehme R., Thomas H.C. (1996). *Viral dynamics in hepatitis B virus infection*, Proceedings of the National Academy of Sciences, USA **93**: 4398-4402.
- [44] Owen A.B. (1994). *Lattice sampling revisited: Monte Carlo variance of means over randomized orthogonal array*, The Annals of Statistics, **22**: 930-945.
- [45] Park J.S., Baek J. (2001). *Efficient computation of maximum likelihood estimators in a spatial linear model with power exponential variogram*, Computer & Geosciences, **27**: 1-7.
- [46] Patterson H.D., Thompson R. (1971). *Recovery of interblock information when block sizes are unequal*, Biometrika, **58**: 545-554.

- [47] Pázman A. (2004). *Correlated optimum design with parametrized covariance function: justification of the Fisher information matrix and of the method of virtual noise*, Research Report Series of the Departmente of Statistics and Mathematics, Wirtschaftsuniversitat Wien, 5.
- [48] Pázman A., Müller W.G. (2000). *Optimal design of experiments subject to correlated errors*, Statistics and Probability Letters, **52**: 29-34.
- [49] Pepelyshev A. (2007). *Optimal Designs for the Exponential Model with Correlated Observations*, López-Fidalgo J., Rodríguez-Díaz J.M. and Torsney B. Eds, mODa 8-Advances in Model-Oriented Design and Analysis, Contributions to Statistics, 165-172, Physica-Verlag.
- [50] Rao C.R. (1973). *Linear Statistical Inference and its Applications*, Wiley, New York.
- [51] Sacks J., Welch W.J., Mitchell T.J., Wynn H.P. (1989). *Design and analysis of computer experiments*, Statistical Science, **4**: 409-435.
- [52] Santner T.J., Williams B.J., Notz W.I. (2003). *The Design and Analysis of Computer Experiments*, Springer, New York.
- [53] Sasena M.J. (2002). *Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations*, PhD thesis, University of Michigan.
- [54] Severini T.A. (2000). *Likelihood Methods in Statistics*, Oxford University Press, Oxford.
- [55] Silvey S.D. (1980). *Optimal Designs*, Chapman & Hall, London.
- [56] Stehlík M. (2007). *D-optimal Designs and Equidistant Designs for Stationary Processes*, López-Fidalgo J., Rodríguez-Díaz J.M. and Torsney B. Eds, mODa 8-Advances in Model-Oriented Design and Analysis, Contributions to Statistics, 165-172, Physica-Verlag.
- [57] Tang B. (1998). *Selecting Latin Hypercubes Using Correlation Criteria*, Statistica Sinica, **8**: 965-978.
- [58] Walter E., Pronzato L. (1997). *Identification of Parametric Models from Experimental Data*, Springer, Paris.

- [59] Zagoraiou M., Baldi Antognini A. (2007). *On the Optimal Designs for Gaussian Ordinary Kriging with exponential correlation structure*, manoscritto sottoposto per la pubblicazione.
- [60] Zhang H., Zimmerman D. (2005). *Towards reconciling two asymptotic frameworks in spatial statistics*, *Biometrika*, **92**: 921-936.
- [61] Zhu Z., Stein M.L. (2005). *Spatial sampling design for parameter estimation of the covariance function*, *Journal of Statistical Planning and Inference*, **134**: 583-603.
- [62] Zimmerman D.L. (2006). *Optimal network for spatial prediction, covariance parameter estimation, and empirical prediction*, *Environmetrics*, **17**: 635-652.