

UNIVERSITÀ DI PISA

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE NATURALI
Corso di Laurea Specialistica in Matematica

**ANALISI STATISTICA DI UN MODELLO
GAUSSIANO PER LO STUDIO
DELL'INFEZIONE DA HIV**



Relatrice:
Prof.ssa Rita Giuliano

Laureando:
Stefano Giampiccolo

Anno Accademico 2010-2011

Indice

1	Introduzione	4
2	Processi stocastici stazionari e terminologia della statistica matematica	8
2.1	I processi stocastici stazionari	8
2.1.1	Prime definizioni	8
2.1.2	Il lemma di Herglotz	12
2.2	Terminologia della statistica matematica	15
2.2.1	Modelli statistici e verosimiglianza	16
2.2.2	Teoria della stima	17
2.2.3	Test statistici e funzioni di test	19
3	I processi gaussiani	21
3.1	Sistemi e processi gaussiani	21
3.2	Filtri lineari	26
3.2.1	Definizioni	26
3.2.2	I processi ARMA(p,q)	29
3.2.3	I processi gaussiani puramente non deterministici	32
3.3	I processi gaussiani di Markov	40
3.3.1	Vettori gaussiani di Markov	40
3.3.2	Processi gaussiani di Markov	45
4	Ricerca della trasformazione per la gaussianità	48
4.1	Le distribuzioni Box Cox	49
4.1.1	Le distribuzioni log-gaussiane	50
4.1.2	Le distribuzioni di Box Cox con $\lambda \neq 0$	53
4.2	Il metodo di Box Cox	59
4.2.1	Costruzione del metodo	59

4.2.2	Un approccio formale e giustificazioni del metodo . . .	62
4.2.3	La scelta della trasformazione	64
4.3	Metodo basato sul test di Kolmogorov-Smirnov	65
4.3.1	Il test di Kolmogorov-Smirnov.	66
4.3.2	Un metodo basato sul test di Kolmogorov Smirnov . .	68
4.3.3	Risultati ottenuti e confronto con il metodo di Box Cox.	76
5	Inferenza statistica sui parametri degli individui sani	78
5.1	Notazione utilizzata	79
5.2	La stima di μ	80
5.2.1	Determinazione degli stimatori di μ	80
5.2.2	È possibile la scelta di uno stimatore migliore per μ ? .	84
5.3	Inferenza statistica sulla funzione di covarianza del processo .	85
5.3.1	La distribuzione di forme quadratiche di vettori gaussiani	86
5.3.2	Stimatori per $\Gamma^*(n)$	91
5.3.3	Inferenza statistica sulla funzione di covarianza nel ca- so markoviano	94
5.4	Risultati ottenuti	96
6	Un approccio alternativo alla stima dei parametri nei sani.	98
6.1	Notazione	99
6.2	La stima della media	99
6.3	La stima della funzione di covarianza: i correlogrammi	102
6.3.1	Condizioni sulla distribuzione spettrale del processo. .	107
6.3.2	Ipotesi particolari sulla struttura del processo	109
6.4	Inferenza statistica sulla funzione di covarianza nel caso il processo sia di Markov	117
6.5	Risultati ottenuti	120
7	Tentativo di convalida del modello	122
7.1	I sieroconvertiti: potenzialità e problemi dei dati	123
7.2	La verifica della trasformazione	125
7.3	Verifica della linearità del processo	125
7.3.1	Il modello lineare generalizzato e lo stimatore di Aitken	126
7.3.2	Distorsione e varianza: il principio di parsimonia	129
7.3.3	La divergenza di Kullback-Leibler	131
7.3.4	Il criterio di informazione secondo Akaiake	133
7.3.5	Correzioni al criterio di informazione di Akaiake	138

7.3.6	Risultati ottenuti	141
7.3.7	Ragioni del fallimento della convalida del modello . . .	144
8	L'analisi dei dati dei sieropositivi	148
8.0.8	La regressione localmente polinomiale	150
8.0.9	Metodo seguito nel nostro caso e risultati ottenuti . . .	153
8.1	La verifica dell'ipotesi di linearità	155
8.2	Il test sulla statistica di Hotelling	157
8.2.1	La distribuzione di Wishart e la statistica di Hotelling.	158
8.2.2	La versione vettoriale del teorema di Cochran	160
8.2.3	Test di ipotesi sulla media di un vettore gaussiano . . .	164
8.3	Risultati ottenuti e possibili modifiche al modello	166
9	Conclusioni	167

Capitolo 1

Introduzione

In questa tesi ci proponiamo di fondare matematicamente l'analisi statistica di un particolare modello gaussiano per lo studio dell'infezione da HIV. Lo scopo non è derivare risultati medici, ma investigare le basi matematiche (probabilistiche) dell'analisi statistica. Discutiamo un modello proposto da C. K. Sleeman e C. J. Mode in [7] per caratterizzare il decadimento del numero di un certo tipo di linfociti ($CD4+$) nel sangue di soggetti sieropositivi.

Il ruolo di questi linfociti nell'infezione da HIV è fondamentale: sebbene l'interazione fra sistema immunitario e virus sia complessa, la progressiva perdita di $CD4+$ è la caratteristica tipica dell'infezione da HIV.

Per modellizzare il numero di $CD4+$ per mm^3 di sangue nei soggetti *sieronegativi*, gli autori si servono di un processo stocastico

$$\{X(t)\}_{t \in \mathbb{R}}.$$

Ipotizzano che questo sia stazionario e log-gaussiano. Per i soggetti *sieropositivi* il numero di $CD4+$ per mm^3 di sangue è modellizzato dal processo stocastico

$$Z(t) = X(t)e^{-\delta t}$$

dove $\delta > 0$ è un parametro di decadimento e $t = 0$ viene fatto corrispondere al momento dell'infezione. Anche il processo $Z(t)$ è log-gaussiano: il relativo processo gaussiano ($\log Z(t)$) ha una funzione delle medie lineare nel tempo,

$$\mathbb{E}\{Z(t)\} = \mu - \delta t,$$

dove μ indica la media, costante, del processo $\log(X(t))$.

Nella tesi mettiamo alla prova le ipotesi avanzate dal modello proposto; quando queste sono accettabili, cerchiamo anche di determinare dei buoni stimatori per i parametri introdotti.

Il capitolo 2 introduce i concetti della statistica matematica (test statistici, modelli e stimatori) e dà alcuni teoremi sui processi stocastici stazionari, di cui studiamo la funzione di covarianza e la distribuzione spettrale. Nel terzo capitolo, invece, ci concentriamo sui processi gaussiani, dimostrando che la *funzione di covarianza* e la *media* di un processo gaussiano stazionario ne caratterizzano le proprietà statistiche. Nello stesso capitolo, presentiamo due particolari tipi di processi gaussiani utili nell'analisi statistica successiva, i processi gaussiani stazionari di Markov e i processi gaussiani stazionari ottenuti dall'applicazione ad un processo stazionario di un filtro lineare invariante nel tempo. I processi del primo tipo hanno una funzione di covarianza parametrizzata; sono l'unico caso di inferenza parametrica sulla funzione di covarianza nella tesi. Per i processi del secondo tipo si possono dimostrare delle proprietà di normalità asintotica per certi stimatori della funzione di covarianza (i *correlogrammi*).

L'analisi statistica vera e propria è condotta a partire dal quarto capitolo. Essa si basa sui dati relativi alla prima coorte del *Public Data Set* del *Multicenter AIDS Cohort Study*: in questa coorte sono presenti soggetti sieronegativi, sieropositivi e sieroconvertiti (individui che sono diventati sieropositivi durante il monitoraggio). Per ogni soggetto sono riportate un massimo di 45 misure, ottenute in visite ripetute ogni sei mesi. Per effettuare i test e per stimare i parametri abbiamo utilizzato il software *R* o scritto degli appositi programmi con il linguaggio *Visual Basic*.

Nel quarto capitolo, sotto l'ipotesi che il processo $\{X(t)\}_{t \in \mathbb{R}}$ sia stazionario, effettuiamo un test sulla sua log-gaussianità ed otteniamo esiti negativi. Cerchiamo allora una trasformazione del processo diversa dal logaritmo e tale che il processo trasformato sia gaussiano. Dapprima utilizziamo il metodo standard in letteratura (metodo di Cox-Box), poi presentiamo un metodo alternativo originale: i due metodi portano a risultati compatibili.

Nel quinto e nel sesto capitolo, sono stimate la funzione di covarianza e la media del processo $\log(X(t))$. L'approccio dei due capitoli è diverso: nel quinto, determiniamo degli stimatori per media e alcuni valori della funzione di covarianza che hanno buone proprietà asintotiche *sul numero di pazienti*, mentre nel sesto capitolo l'asintoticità è intesa *sul numero di visite* effettua-

te dal singolo paziente. I risultati ottenuti dai due approcci sono diversi: nel capitolo 5, sotto l'ipotesi dell'indipendenza fra le misure relative a pazienti diversi, dimostriamo che le successioni di stimatori considerate sono successioni fortemente consistenti, nel capitolo successivo invece i risultati asintotici sugli stimatori sono ricavati facendo alcune ipotesi sul processo $\{\log(X(t))\}_{t \in \mathbb{R}}$. Visto che i risultati asintotici del capitolo 6 sono comunque più deboli, nel seguito utilizziamo soltanto le stime del quinto capitolo.

Infine, nei capitoli 7 ed 8 ci proponiamo di sottoporre a test l'ipotesi della linearità della funzione delle medie del processo $\{\log Z(t)\}_{t \in \mathbb{R}^+}$, ipotizzando che la funzione di covarianza del processo $\{\log Z(t)\}_{t \in \mathbb{R}^+}$ coincida con quella del processo $\{\log X(t)\}_{t \in \mathbb{R}}$. Nel settimo capitolo consideriamo i dati dei *sieroconvertiti*: essi sono particolarmente importanti, perché per loro è possibile datare con una buona approssimazione l'infezione da HIV. Il modello lineare *non* sembra essere adeguato, probabilmente a causa del periodo di latenza del virus HIV e dell'infezione primaria - due fenomeni trascurati nel modello che consideriamo. Cerchiamo allora di verificare se il modello è valido almeno per il periodo *successivo* a quello di latenza, attraverso i dati relativi ai pazienti sieropositivi (capitolo 8). Non è possibile considerare questi dati tutti insieme, perché non c'è alcun indizio sulla data dell'infezione dei diversi pazienti. Per isolare i soggetti che manifestano un trend decrescente del numero di linfociti $CD4+$ operiamo una regressione non parametrica sulle misure di ogni individuo, selezionando in questo modo solo quelli che, presumibilmente, stanno attraversando il periodo successivo a quello della latenza del virus. Testiamo l'ipotesi di linearità della funzione delle medie del processo $\{\log Z(t)\}_{t \in \mathbb{R}^+}$ con il test di Hotelling; l'ipotesi non sembra confermata.

Nella tesi, sono proposte alcune modifiche al modello proposto in [7]: riguardo alla scelta della trasformazione T che rende il processo $\{T(X(t))\}_{t \in \mathbb{R}^+}$ gaussiano, al ruolo del periodo di latenza, e all'ipotesi (che riteniamo si possa scartare) di un decadimento lineare del numero (opportunamente trasformato) di linfociti $CD4+$ per mm^3 di sangue in una persona sieropositiva. Per quanto riguarda la trasformazione, è importante notare che quella determinata nella tesi *non* è universale, tanto che altri campioni, diversi da quello utilizzato da noi, fanno pensare ad una trasformazione completamente diversa.

Affrontando il problema della ricerca della trasformazione, nel quarto capitolo, abbiamo proposto un metodo originale, alternativo a quello di Cox-

Box. Data una variabile aleatoria X , quasi certamente positiva, supponiamo che esista $\alpha \in [0, 1]$ tale che X^α abbia una particolare distribuzione \mathbb{P}_0 . Ponendo alcune ipotesi su X e su \mathbb{P}_0 , abbiamo determinato una successione di stimatori fortemente consistente di α , basata sulle osservazioni di realizzazioni indipendenti di X .

Capitolo 2

Processi stocastici stazionari e terminologia della statistica matematica

Questo capitolo è introduttivo e si divide in due parti.

Nella prima parte (sezione 2.1), diamo le definizioni di funzione delle medie e funzione di covarianza di un processo stocastico. Caratterizziamo poi l'insieme delle funzioni di covarianza grazie al Lemma di Herglotz, che dimostra l'esistenza della distribuzione spettrale di un processo stazionario. Questi concetti saranno utilizzati largamente nei capitoli 5 e 6, per la stima dei parametri degli individui sieronegativi.

La seconda parte (sezione 2.2), invece, introduce i concetti di test statistici, stimatori, campioni. . . , che serviranno in tutta la tesi.

2.1 I processi stocastici stazionari

2.1.1 Prime definizioni

Definizione 2.1.1. *Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e $T \subset \mathbb{R}$. Una famiglia di variabili aleatorie $\{X(t)\}_{t \in T}$ definite su $(\Omega, \mathcal{F}, \mathbb{P})$ e a valori in (E, \mathcal{E}) si dice un processo stocastico definito su $(\Omega, \mathcal{F}, \mathbb{P})$ a valori in (E, \mathcal{E}) . Quando si può sottointendere senza ambiguità la tribù \mathcal{E} , come nel caso degli spazi $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ o $(\mathbb{C}^n, \mathcal{B}(\mathbb{C}^n))$, si parlerà semplicemente di processo stocastico a valori in E .*

Un processo si dice a *tempi discreti* se $T = \mathbb{N}$ o $T = \mathbb{Z}$, a *tempi continui* se $T \subset \mathbb{R}$ è un intervallo, una semiretta o tutto \mathbb{R} .

Sia $\{X(t)\}_{t \in T}$ un processo stocastico a valori complessi. Diremo che il processo è *regolare* se

$$\mathbb{E}\{|X(t)|^2\} < \infty \quad \forall t \in T.$$

In questo caso ha senso definire la *funzione delle medie* del processo come

$$\begin{aligned} \mu &: T \rightarrow \mathbb{C} \\ t &\mapsto \mathbb{E}\{X(t)\} \end{aligned}$$

e la *funzione di covarianza*

$$\Gamma : T \times T \rightarrow \mathbb{C}$$

$$(t, t') \mapsto \text{cov}\{X(t), X(t')\} = \mathbb{E}\{(X(t) - \mu(t))\overline{(X(t') - \mu(t'))}\}.$$

La funzione di covarianza quindi associa a ciascuna coppia (t, t') la covarianza delle variabili $X(t)$ e $X(t')$.

Definiamo le funzioni semi-definite positive, che serviranno per caratterizzare le funzioni di covarianza di un processo.

Definizione 2.1.2. *Sia A un insieme e $g : A \times A \rightarrow \mathbb{C}$, si dice che g è semidefinita positiva se*

1. g è "hermitiana", cioè $g(a, a') = \overline{g(a', a)}$ per ogni $a, a' \in A$;
2. per ogni $A_n = \{a_1, \dots, a_n\} \subseteq A$ e per ogni $h : A_n \rightarrow \mathbb{C}$ si ha

$$\sum_{a, a' \in A_n} g(a, a') h(a) \overline{h(a')} \geq 0.$$

g si dice definita positiva se è semidefinita positiva e se, per ogni $A_n = \{a_1, \dots, a_n\} \subseteq A$ e per ogni $h : A_n \rightarrow \mathbb{C}$

$$\sum_{a, a' \in A_n} g(a, a') h(a) \overline{h(a')} = 0$$

se e solo se $h(a) = 0$ per ogni $a \in A_n$.

Proposizione 2.1.1. *Supponiamo che $T \subset \mathbb{R}$ e sia F una funzione $F : T \times T \rightarrow \mathbb{C}$. F è la funzione di covarianza di un processo stocastico $\{X(t)\}_{t \in T}$ se e solo se è semidefinita positiva.*

Dimostrazione. Dimostriamo solo che la funzione di covarianza Γ di un processo ha queste proprietà, l'altra implicazione verrà dimostrata con la proposizione 3.1.1. Supponiamo per semplicità che il processo sia centrato, cioè che $\mathbb{E}\{X(t)\} = 0$ per ogni $t \in T$. Dato che

$$\overline{\Gamma(t, t')} = \overline{\mathbb{E}\{X(t)\overline{X(t')}\}} = \mathbb{E}\{X(t')\overline{X(t)}\},$$

Γ è una funzione hermitiana. Inoltre, assegnati $T_n = \{t_1, \dots, t_n\} \subseteq T$ e $h : T_n \rightarrow \mathbb{C}$

$$\begin{aligned} \sum_{t, t' \in T_n} \Gamma(t, t') h(t) \overline{h(t')} &= \sum_{t, t' \in T_n} \mathbb{E}\{X(t)\overline{X(t')}\} h(t) \overline{h(t')} = \\ &= \mathbb{E}\left\{ \sum_{t \in T_n} X(t) h(t) \overline{\sum_{t' \in T_n} X(t') h(t')} \right\} = \mathbb{E}\left\{ \left\| \sum_{t \in T_n} h(t) X(t) \right\|^2 \right\} \geq 0. \end{aligned}$$

Dunque la funzione di covarianza di un processo è semidefinita positiva. \square

Nei prossimi capitoli saremo interessati principalmente ai processi stocastici stazionari, che ora definiamo.

Definizione 2.1.3. *Un processo stocastico $\{X(t)\}_{t \in T}$, regolare a valori complessi si dice stazionario in senso lato, o debolmente stazionario, se la funzione delle medie $\mu(t)$ è costante e, per ogni $t, t' \in T$ e $h \in \mathbb{R}$ tali che $t + h, t' + h \in T$, si ha*

$$\Gamma(t, t') = \Gamma(t + h, t' + h).$$

La funzione di covarianza dipende cioè esclusivamente dalla differenza $t - t'$.

Si può dare un'altra definizione di stazionarietà, basata sulle distribuzioni di probabilità delle variabili del processo.

Definizione 2.1.4. *Un processo stocastico $\{X(t)\}_{t \in T}$ si dice stazionario in senso stretto se, per ogni $n \in \mathbb{N}$, per ogni $h \in \mathbb{R}$ e per ogni scelta di $\{t_1, \dots, t_n\} \subset T$ tale che $\{t_1 + h, \dots, t_n + h\} \subset T$, i vettori aleatori*

$$(X(t_1), \dots, X(t_n)) \quad (X(t_1 + h), \dots, X(t_n + h))$$

hanno la stessa legge.

Naturalmente se un processo è stazionario in senso stretto, lo è anche in senso lato. Come osservato sopra, se un processo è stazionario in senso lato, la sua funzione di covarianza è in realtà una funzione della differenza $t - t'$. Formalizziamo la situazione in questo modo: sia $S \subseteq \mathbb{R}$, indichiamo con D_S l'insieme

$$D_S = \{u - v \mid u, v \in S\}.$$

Ovviamente $0 \in D_S$ e l'insieme D_S è simmetrico, cioè se $u \in D_S$ allora $-u \in D_S$. D_S è isomorfo al quoziente di S rispetto alla relazione di equivalenza in $S \times S$ definita da

$$(a, b) \sim (a', b') \text{ se e solo se } \exists h \quad (a + h, b + h) = (a', b').$$

Se il processo è stazionario in senso lato, dunque, la funzione di covarianza passa al quoziente, definendo una funzione

$$g : D_S \rightarrow \mathbb{C}$$

$$g(u - v) = \Gamma(u, v).$$

D'ora in poi chiameremo anche g *funzione di covarianza* del processo; adattiamo per questo la definizione di funzione semidefinita positiva data in precedenza.

Definizione 2.1.5. *Sia $S \subseteq \mathbb{R}$ e $g : D_S \rightarrow \mathbb{C}$. g è semidefinita positiva se*

1. g è "hermitiana" cioè $g(u) = \overline{g(-u)}$.
2. per ogni $S_n = \{d_1, \dots, d_n\} \subseteq S$ e $h : S_n \rightarrow \mathbb{C}$ si ha

$$\sum_{d, d' \in S_n} g(d - d') h(d) \overline{h(d')} \geq 0.$$

La funzione g è semidefinita positiva se e solo lo è il suo sollevamento a $D \times D$. Dunque la funzione di covarianza di un processo $\{X(t)\}_{t \in T}$ stazionario in senso lato, intesa come funzione con dominio D_T , è semidefinita positiva. Nella proposizione 3.1.1, inoltre, proveremo che se g è una funzione semidefinita positiva su D_T allora esiste un processo stocastico $\{X(t)\}_{t \in T}$ la cui funzione di covarianza è proprio g . Lo studio delle funzioni di covarianza può essere perciò ricondotto allo studio delle funzioni semidefinite positive. Nei prossimi capitoli considereremo principalmente processi a tempi discreti in cui $T = \mathbb{N}$ o $T = \mathbb{Z}$. In questi casi $D_T = \mathbb{Z}$.

2.1.2 Il lemma di Herglotz

In questo paragrafo dimostreremo l'esistenza della distribuzione di probabilità spettrale di un processo a tempi discreti e stazionario in senso lato, che verrà utilizzata nei capitoli successivi.

Proposizione 2.1.2. *Supponiamo che $S \subset \mathbb{R}$ e che g sia una funzione semidefinita positiva su D_S , allora:*

1. $g(0) \geq 0$.
2. per ogni $u \in D_S$ si ha $|g(u)| \leq g(0)$.

Dimostrazione. Per provare il primo punto basta applicare la definizione di funzione semidefinita positiva con $S = \{u\}$ e $h(u) = 1$.

Per quanto riguarda il secondo asserto, consideriamo l'insieme

$$S = \{x_1, \dots, x_n\}.$$

La definizione di funzione semidefinita positiva implica che la matrice

$$A = \{a_{i,j}\}_{i,j=1,\dots,n}$$

definita da

$$a_{i,j} = g(x_i - x_j)$$

sia semidefinita positiva; il suo determinante deve essere dunque maggiore o uguale a 0. In particolare prendiamo l'insieme $S = \{x, x + u\}$, con $x \in \mathbb{R}$; allora la condizione di positività del determinante diventa

$$g(0)^2 \geq g(u)^2,$$

condizione che, grazie al punto 1, può essere riscritta come

$$g(0) \geq |g(u)|.$$

□

La proposizione che segue è nota come *lemma di Herglotz*.

Proposizione 2.1.3. *Sia g una funzione*

$$g : \mathbb{Z} \rightarrow \mathbb{C},$$

essa è semidefinita positiva se e solo se esiste una funzione F crescente in senso lato, continua a destra, limitata, con $F(-\pi) = 0$ tale che per ogni $h \in \mathbb{Z}$

$$g(h) = \int_{-\pi}^{\pi} e^{ihx} dF(x),$$

dove dF indica l'integrale di Lebesgue-Stieltjes.

Dimostrazione. Indichiamo con E l'intervallo

$$E = (-\pi, \pi].$$

Se

$$g(h) = \int_E e^{ihx} dF(x), \quad h \in \mathbb{Z}$$

allora g è la restrizione a \mathbb{Z} della funzione caratteristica della misura finita che ha come funzione di ripartizione $G = I_E F$, dove I_E è l'indicatrice di E . Per le note proprietà delle funzioni caratteristiche g è hermitiana, inoltre per ogni $u, v \in \mathbb{Z}$ e per ogni $h : \mathbb{Z} \rightarrow \mathbb{C}$:

$$\begin{aligned} \sum_{u,v} g(u-v)h(u)\overline{h(v)} &= \int_{\mathbb{R}} \sum_{u,v} e^{i(u-v)x} h(u)\overline{h(v)} dG(x) = \\ &= \int_E \left\| \sum_u e^{iux} h(u) \right\|^2 dG(x) \geq 0. \end{aligned}$$

Dimostriamo l'implicazione opposta: dato che g è semidefinita positiva, abbiamo

$$\begin{aligned} f_n(x) &:= \frac{1}{2\pi n} \sum_{j=1}^n \sum_{h=1}^n g(j-h) e^{-i(j-h)x} = \\ &= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n}\right) g(k) e^{-ikx} \geq 0. \end{aligned}$$

L'insieme di funzioni $\{e^{ikx} | k \in \mathbb{N}\}$ è un sistema ortonormale di $L^2(E, \mathcal{B}(E), \lambda)$, dove λ è la misura di Lebesgue.

Quindi, integrando sull'insieme E , per ogni k con $|k| < n$ troviamo

$$\left(1 - \frac{|k|}{n}\right) g(k) = \int_E e^{ikx} f_n(x) dx = \int_E e^{ikx} dF_n(x), \quad (2.1)$$

dove abbiamo indicato con F_n la funzione di ripartizione della misura finita avente densità, rispetto alla misura di Lebesgue,

$$f_n(x) \cdot I_E.$$

La successione di misure definite dalle F_n è tesa, perché, per (2.1),

$$F_n(\pi) = \int_E dF_n(x) = g(0).$$

Dunque per il teorema di relativa compattezza di Helly esiste una sottosuccessione $\{F_{n_k}\}_{k \in \mathbb{N}}$ che converge strettamente ad una distribuzione finita F . Sia $f : \mathbb{R} \rightarrow \mathbb{C}$ limitata e continua, secondo la definizione di convergenza stretta:

$$\lim_{k \rightarrow \infty} \int f dF_{n_k} = \int f dF.$$

Scegliamo $f(x) = e^{iux}$: passando al limite sulla sottosuccessione $\{F_{n_k}\}_{k \in \mathbb{N}}$ otteniamo, dalla (2.1),

$$g(u) = \lim_k \left(1 - \frac{|u|}{n_k}\right) g(u) = \lim_k \int_E e^{iux} dF_{n_k}(x) = \int_E e^{iux} dF(x).$$

□

Dunque possiamo dare la seguente caratterizzazione delle funzioni di covarianza di un processo stocastico stazionario in senso lato e a tempi discreti.

Corollario 2.1.1. *Sia g funzione $g : \mathbb{Z} \rightarrow \mathbb{C}$. Le seguenti affermazioni sono equivalenti:*

1. g è la funzione di covarianza di un processo stazionario in senso lato $\{X(n)\}_{n \in T}$, con $T = \mathbb{N}$ o $T = \mathbb{Z}$.
2. $g(h) = \int_{(-\pi, \pi]} e^{iht} dF(t)$ per ogni $h \in \mathbb{Z}$, dove F è la funzione di ripartizione di una misura finita concentrata su $(-\pi, \pi]$.
3. g è una funzione semidefinita positiva.

Dimostrazione. è una semplice applicazione del lemma di Herglotz e di quanto provato a proposito dei processi stazionari in senso lato. □

La funzione di ripartizione F si dice *funzione di ripartizione spettrale*. Se essa è assolutamente continua rispetto alla misura di Lebesgue, la sua densità f si dice *densità spettrale* della funzione g o equivalentemente del processo $\{X(n)\}_{n \in \mathbb{N}}$ di cui g è funzione di covarianza.

In generale la misura indotta da F è semplicemente una misura finita: essa è la misura nulla se e solo se $g(0) = 0$, in caso contrario ci si può ricondurre ad una distribuzione di probabilità spettrale normalizzando la misura spettrale. Vediamo come si possano semplificare i risultati ottenuti se ipotizziamo che la funzione g sia a valori reali. Per quanto detto possiamo limitarci a considerare funzioni semidefinite positive g tali che $g(0) = 1$. Per le note proprietà delle funzioni caratteristiche, la funzione g è reale se e solo se la distribuzione di probabilità spettrale è simmetrica rispetto allo 0, cioè se e solo se, per ogni $\lambda \in (-\pi, \pi)$,

$$F(\lambda) = F(\pi^-) - F(-\lambda^-).$$

In questo caso dunque si può semplificare l'espressione di g come segue

$$g(h) = \int_{(-\pi, \pi]} e^{iht} dF(t) = \int_{(-\pi, \pi]} \cos(ht) dF(t),$$

e, una volta definita univocamente la misura di probabilità G su $(0, \pi]$ che soddisfa le seguenti condizioni:

$$G(\lambda) = 2(F(\lambda) - F(0)) + F(0) - F(0^-) \text{ se } 0 \leq \lambda < \pi$$

$$G(\pi) - G(\pi^-) = F(\pi),$$

l'espressione di g può essere ulteriormente semplificata, perché

$$g(h) = \int_{(-\pi, \pi]} \cos(ht) dF(t) = \int_{(0, \pi]} \cos(ht) dG(t).$$

2.2 Terminologia della statistica matematica

In questa sezione vogliamo introdurre e spiegare concisamente la terminologia della statistica matematica che utilizzeremo nei capitoli successivi.

2.2.1 Modelli statistici e verosimiglianza

Definizione 2.2.1. Sia (Ω, \mathcal{F}) uno spazio misurabile e I un insieme di misure di probabilità su di esso. La terna

$$(\Omega, \mathcal{F}, I)$$

si dice modello statistico.

Nel caso esista una parametrizzazione iniettiva dell'insieme di probabilità I

$$I = \{\mathbb{P}^\theta\}_{\theta \in \Theta},$$

con $\Theta \subset \mathbb{R}^n$, la terna

$$(\Omega, \mathcal{F}, \{\mathbb{P}^\theta\}_{\theta \in \Theta})$$

si dice modello statistico parametrico.

Supponiamo di avere un modello statistico parametrico $(\Omega, \mathcal{F}, \{\mathbb{P}^\theta\}_{\theta \in \Theta})$ e una misura di probabilità \mathbb{Q} tale che ogni \mathbb{P}^θ sia assolutamente continua rispetto alla misura \mathbb{Q} . Se L è una funzione

$$L : \Theta \times \Omega \mapsto \mathbb{R}$$

tale che, per ogni θ , la variabile aleatoria

$$\omega \mapsto L(\theta, \omega)$$

è una versione della densità di \mathbb{P}^θ rispetto a \mathbb{Q} , L si dice *verosimiglianza* del modello statistico.

Altra nozione fondamentale della statistica matematica è quella di campione.

Definizione 2.2.2. Siano $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità, (E, \mathcal{E}) uno spazio misurabile e $Y_1, \dots, Y_n : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ una n -pla di variabili aleatorie indipendenti. Il vettore aleatorio

$$(Y_1, \dots, Y_n)$$

si dice un campione di taglia n a valori in E .

Se, nella definizione 2.2.2, anziché avere uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ avessimo un modello statistico $(\Omega, \mathcal{F}, \{\mathbb{P}^\theta\}_{\theta \in \Theta})$, si può estendere la definizione di campione.

Supponiamo che per ogni misura di probabilità \mathbb{P}^θ , le variabili aleatorie

$$Y_1, \dots, Y_n$$

siano indipendenti, allora il vettore aleatorio (Y_1, \dots, Y_n) si dice *campione*. Se Y_1, \dots, Y_n sono identicamente distribuite, per ogni \mathbb{P}^θ , si dice che il campione è a variabili *identicamente distribuite con legge* \mathbb{P}^θ , dove $\mathbb{P}_{Y_i}^\theta$ indica la probabilità immagine di \mathbb{P}^θ secondo la variabile aleatoria Y_i .

Continuiamo a supporre che nella definizione 2.2.2 si abbia un modello statistico

$$(\Omega, \mathcal{F}, \{\mathbb{P}^\theta\}_{\theta \in \Theta}),$$

possiamo dotare anche lo spazio (E^n, \mathcal{E}^n) di una famiglia di probabilità parametrizzata da θ , considerando le distribuzioni \mathbb{P}_Y^θ , dove \mathbb{P}_Y^θ indica la probabilità immagine di \mathbb{P}^θ attraverso il vettore aleatorio

$$Y = (Y_1, \dots, Y_n).$$

Il modello statistico

$$(E^n, \mathcal{E}^n, \{\mathbb{P}_Y^\theta\}_{\theta \in \Theta})$$

si dice modello statistico *indotto* dal campione Y_1, \dots, Y_n .

2.2.2 Teoria della stima

Definizione 2.2.3. Sia $(\Omega, \mathcal{F}, \{\mathbb{P}^\theta\}_{\theta \in \Theta})$ un modello statistico parametrico e $f : \Theta \mapsto \mathbb{R}^k$. Una funzione misurabile che non dipende da θ

$$X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$$

si dice *stimatore di $f(\theta)$* .

Se per ogni $\theta \in \Theta$ la variabile aleatoria X della definizione precedente è integrabile e, per ogni θ

$$\mathbb{E}^\theta\{X\} = f(\theta),$$

X si dice uno stimatore *corretto*, o *non distorto*, di $f(\theta)$. Supponiamo poi di avere una successione $\{X_n\}_{n \in \mathbb{N}}$ di stimatori di $f(\theta)$. Se per ogni θ la convergenza

$$X_n \rightarrow f(\theta)$$

ha luogo in qualche senso, la successione di stimatori di $f(\theta)$ si dice *consistente*. La successione si dice *fortemente consistente* se per ogni θ si ha

$$\lim_n X_n = f(\theta) \quad \mathbb{P}^\theta - \text{q.c.}$$

La definizione di stimatori data è molto generica: non ci dice in alcun modo “quanto bene” stiamo stimando $f(\theta)$. Introduciamo la definizione di costo.

Definizione 2.2.4. Sia $(\Omega, \mathcal{F}, \{\mathbb{P}^\theta\}_{\theta \in \Theta})$ un modello statistico, $f : \Theta \rightarrow \mathbb{R}^k$ e X uno stimatore di $f(\theta)$. Assegnata una funzione misurabile $C : \mathbb{R}^k \times \mathbb{R}^k \mapsto \mathbb{R}$ a valori positivi, la funzione

$$\omega \mapsto C(f(\theta), X(\omega))$$

si dice costo della sostituzione di $f(\theta)$. Dove non vi è rischio di creare ambiguità la chiameremo semplicemente costo.

Nei prossimi capitoli noi utilizzeremo solo il costo quadratico, che si ottiene con la funzione

$$\begin{aligned} C : \mathbb{R}^k \times \mathbb{R}^k &\rightarrow \mathbb{R} \\ (x, y) &\mapsto \|x - y\|^2. \end{aligned}$$

Il costo, in conseguenza della sua definizione, è una variabile aleatoria. Se supponiamo che essa sia integrabile per ogni probabilità \mathbb{P}^θ , possiamo definire il *costo medio* dello stimatore X , come la funzione

$$\theta \mapsto \mathbb{E}^\theta\{C(f(\theta), X(\omega))\}.$$

Nel caso che considereremo dunque, il costo medio sarà indicato col termine *rischio quadratico medio* dello stimatore X , definito da

$$\theta \mapsto \mathbb{E}^\theta\{\|X - f(\theta)\|^2\}.$$

Nel caso lo stimatore sia corretto, il rischio quadratico medio è la varianza di X . Gli stimatori saranno confrontati utilizzando il costo medio, secondo la seguente definizione.

Definizione 2.2.5. Nelle ipotesi specificate nella definizione precedente, uno stimatore X si dice preferibile ad uno stimatore Y di $f(\theta)$ se per ogni $\theta \in \Theta$

$$\mathbb{E}^\theta\{C(f(\theta), X)\} \leq \mathbb{E}^\theta\{C(f(\theta), Y)\}.$$

Assegnato un insieme di stimatori $\{X_i\}_{i \in I}$ e un suo elemento X , X si dice uno stimatore ottimale nella classe $\{X_i\}_{i \in I}$ se X è preferibile ad ogni altro elemento di $\{X_i\}_{i \in I}$.

Ci sono delle situazioni in cui non è necessaria una stima puntuale dei parametri, ma è sufficiente specificare un insieme a cui il parametro appartiene con una certa probabilità: si definiscono per questo le regioni di confidenza.

Definizione 2.2.6. *Assegnato un modello statistico $(\Omega, \mathcal{F}, \{\mathbb{P}^\theta\}_{\theta \in \Theta})$, siano $\alpha \in (0, 1)$ e*

$$S : \Omega \rightarrow \mathcal{P}(\Theta)$$

una funzione tale che per ogni $\theta \in \Theta$, l'insieme $\{\omega \in \Omega : \theta \notin S(\omega)\}$ appartenga alla tribù \mathcal{F} . Si dice che S è una regione di fiducia di livello $1 - \alpha$ se per ogni $\theta \in \Theta$, risulta:

$$\mathbb{P}^\theta(\theta \notin S) \leq \alpha.$$

2.2.3 Test statistici e funzioni di test

Nei prossimi capitoli utilizzeremo più volte i test statistici: in questo caso non useremo nelle definizioni modelli parametrici, perché molti dei test utilizzati sono non parametrici, ossia basati su modelli non parametrici.

Definizione 2.2.7. *Sia (Ω, \mathcal{F}, I) un modello statistico e (H_0, H_1) una partizione dell'insieme I . Nel contesto della teoria dei test statistici H_0 si dice ipotesi nulla, mentre H_1 si dice alternativa.*

Supponiamo di avere specificato un'ipotesi nulla H_0 ed un'alternativa H_1 . Sia $A \subset \Omega$ un insieme \mathcal{F} -misurabile: la funzione I_A , indicatrice di A , si dice *funzione di test* di H_0 contro H_1 , o più genericamente test di H_0 contro H_1 e A si dice *regione di rifiuto*.

Anche in questo caso, come accadeva per gli stimatori, la definizione di test statistico non ci dice nulla riguardo all'effettiva efficacia del test. Per questo introduciamo la taglia di un test.

Definizione 2.2.8. *Sia (Ω, \mathcal{F}, I) un modello statistico, I_A una funzione di test dell'ipotesi nulla H_0 contro l'alternativa H_1 .*

- *Si chiama livello di significatività o taglia del test la quantità*

$$\alpha^* := \sup_{\mathbb{P} \in H_0} \mathbb{P}(I_A = 1).$$

- *Assegnato $\alpha \in (0, 1)$, si dice che il test ha livello α se*

$$\alpha^* \leq \alpha.$$

Dunque se un test ha livello α , la probabilità di cadere nella regione di rigetto quando H_0 è vera è minore di α .

Definizione 2.2.9. Sia (Ω, \mathcal{F}, I) un modello statistico e I_{A_β} , con $\beta \in [0, 1]$, una famiglia di test statistici definiti su di esso; supponiamo che le regioni di rifiuto dei test siano nested, cioè che, se $\beta_1 < \beta_2$, si abbia

$$A_{\beta_1} \subset A_{\beta_2}.$$

Allora la variabile aleatoria

$$p : (\Omega, \mathcal{F}) \rightarrow ([0, 1], \mathcal{B}[0, 1])$$

$$\omega \mapsto \inf_{\beta} \{\beta | I_{A_\beta}(\omega) = 1\}$$

si dice p-value della famiglia di test I_{A_β} .

Supponiamo infine di avere un modello statistico parametrico $(\Omega, \mathcal{F}, \{\mathbb{P}^\theta\}_{\theta \in \Theta})$ e che, per ogni $\theta_0 \in \Theta$, esista un test dell'ipotesi $H_0 : \theta = \theta_0$ contro l'alternativa $H_1 : \theta \neq \theta_0$ di livello α ; indichiamo con $D(\theta_0)$ la sua regione di rifiuto. Possiamo costruire una regione di fiducia di livello $1 - \alpha$ per il parametro θ in questo modo:

$$S : \Omega \rightarrow \mathcal{P}(\Theta)$$

$$\omega \mapsto \{\theta \in \Theta : \omega \notin D(\theta)\}.$$

Capitolo 3

I processi gaussiani

In questo capitolo studiamo alcune proprietà dei processi gaussiani, sui quali si basa il modello per l'infezione da HIV proposto in [7]. Dimostriamo che, in generale, la funzione di covarianza e la media di un processo gaussiano stazionario ne caratterizzano le proprietà statistiche. Questo risultato limita l'inferenza statistica sviluppata nei capitoli 5 e 6 all'inferenza su media e funzione di covarianza.

Introduciamo poi dei casi particolari di processi gaussiani: i processi gaussiani di Markov e i processi gaussiani ottenuti tramite l'applicazione ad un altro processo di un filtro lineare invariante nel tempo ad un altro processo.

Dimostriamo che ipotizzare che un processo gaussiano sia di Markov equivale a parametrizzare la sua funzione di covarianza: in questo caso, dunque, è possibile procedere con metodi statistici parametrici per la stima della funzione di covarianza.

Il secondo tipo di processi gaussiani è utilizzato nel capitolo 5 per dimostrare le proprietà asintotiche dei correlogrammi.

3.1 Sistemi e processi gaussiani

In questo paragrafo vogliamo spiegare che cosa sono i sistemi e i processi gaussiani e dimostrare alcuni risultati che saranno utili nei paragrafi successivi.

Definizione 3.1.1. *Sia A un insieme e*

$$\{X(a)|a \in A\}$$

una famiglia di variabili aleatorie definite su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Essa si dice un sistema gaussiano se, per ogni $n \in \mathbb{N}$ e per ogni $\{a_1, \dots, a_n\} \subseteq A$, il vettore aleatorio

$$(X(a_1), \dots, X(a_n))$$

è un vettore gaussiano.

Se un processo stocastico $\{X(t)\}_{t \in T}$ è un sistema gaussiano, esso si dice processo gaussiano.

Per dimostrare che i processi gaussiani effettivamente esistono sfruttiamo il teorema di consistenza di Kolmogorov, che enunciamo di seguito (per una dimostrazione, vedere ad esempio [27, p. 159]).

Teorema 3.1.1. *Sia $T \subseteq \mathbb{R}$ e supponiamo che per ogni $k \in \mathbb{N}$ ed ogni k -upla (t_1, \dots, t_k) di elementi di T , sia assegnata una misura di probabilità $\mu_{(t_1, \dots, t_k)}$ su \mathbb{R}^n , in modo che siano soddisfatte le seguenti ipotesi:*

1. *Se π è una permutazione dell'insieme $\{1, \dots, k\}$ e $F_1, \dots, F_k \subseteq \mathbb{R}$ sono insiemi misurabili, allora:*

$$\mu_{(\pi(t_1), \dots, \pi(t_k))}(F_1 \times \dots \times F_k) = \mu_{(t_1, \dots, t_k)}(F_{\pi^{-1}(1)} \times \dots \times F_{\pi^{-1}(k)}).$$

2. *Se $F_1, \dots, F_k \subseteq \mathbb{R}$ sono insiemi misurabili, $m \in \mathbb{N}$ e $(t_{k+1}, \dots, t_{k+m})$ è una m -upla di elementi di T , si ha:*

$$\mu_{(t_1, \dots, t_k)}(F_1 \times \dots \times F_k) = \mu_{(t_1, \dots, t_k, t_{k+1}, \dots, t_{k+m})}(F_1 \times \dots \times F_k \times \mathbb{R} \times \dots \times \mathbb{R}).$$

Allora esistono uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e un processo stocastico $\{X(t)\}_{t \in T}$ definito su di esso che, per ogni k , soddisfa:

$$\mu_{(t_1, \dots, t_k)}(F_1, \dots, F_k) = \mathbb{P}\{X_{t_1} \in F_1, \dots, X_{t_k} \in F_k\},$$

per ogni scelta di $\{t_1, \dots, t_k\} \subseteq T$ e di $F_1, \dots, F_k \subseteq \mathbb{R}$ misurabili.

Alla dimostrazione dell'esistenza di un processo gaussiano facciamo precedere anche un lemma sui vettori gaussiani.

Lemma 3.1.1. *Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e*

$$X = (X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$$

un vettore gaussiano. Denotiamo con $A = \{a_{i,j}\}_{i,j=1, \dots, n}$ la sua matrice di covarianza e con $\mathbf{b} = (b_1, \dots, b_n)$ il vettore delle medie. Valgono le seguenti affermazioni:

- Se π è una permutazione dell'insieme $\{1, \dots, n\}$, il vettore aleatorio

$$\pi(X) = (X_{\pi(1)}, \dots, X_{\pi(n)})$$

è ancora gaussiano, con vettore delle medie

$$\pi(\mathbf{b}) = (b_{\pi(1)}, \dots, b_{\pi(n)})$$

e matrice di covarianza $\pi(A)_{i,j} = a_{\pi^{-1}(i), \pi^{-1}(j)}$.

- Se $m \leq n$, il vettore (X_1, \dots, X_m) è ancora gaussiano, con vettore delle medie (b_1, \dots, b_m) e matrice di covarianza $\{a_{i,j}\}_{i,j=1,\dots,m}$.

Dimostrazione. La seconda affermazione è ovvia, dimostriamo solo la prima. $\pi(X)$ è ovviamente un vettore gaussiano, inoltre esiste un operatore vettoriale $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tale che:

$$T(X) = \pi(X).$$

Se e_1, \dots, e_n è la base canonica di \mathbb{R}^n , formata da vettori colonna, la matrice associata a T è

$$M_T = [e_{\pi(1)} \quad e_{\pi(2)} \quad \dots \quad e_{\pi(n)}].$$

Dunque la matrice di covarianza di $\pi(X)$ è:

$$M_T A M_T^t = \pi(A).$$

In maniera del tutto analoga si dimostra che $\mathbb{E}\{T(X)\} = \pi(\mathbf{b})$. □

Adesso possiamo dimostrare l'esistenza dei processi gaussiani.

Proposizione 3.1.1. *Sia $T \subseteq \mathbb{R}$, $\Xi : T \times T \rightarrow \mathbb{R}$ una funzione semidefinita positiva e $m : T \rightarrow \mathbb{R}$ una funzione qualsiasi. Esiste uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e un processo stocastico $\{X(t)\}_{t \in T}$ definito su di esso che abbia m come funzione delle medie e Ξ come funzione di covarianza.*

Dimostrazione. Consideriamo per ogni $n \in \mathbb{N}$ ed ogni n -upla (t_1, \dots, t_n) , con $t_i \in T$ per ogni i , la matrice $Q_{(t_1, \dots, t_n)}$ ed il vettore $b_{(t_1, \dots, t_n)}$ così definiti:

$$\{q_{(t_1, \dots, t_n)}\}_{i,j=1,\dots,n} = \Xi(t_i, t_j) \quad b_{(t_1, \dots, t_n)} = (m(t_1), \dots, m(t_n)).$$

Denotiamo con $N_n(\mathbf{b}, Q)$ la distribuzione gaussiana n -dimensionale con vettore delle medie \mathbf{b} e matrice di covarianza Q : ad ogni n -upla (t_1, \dots, t_n) corrisponde dunque una distribuzione di probabilità $N_n(Q_{(t_1, \dots, t_n)}, b_{(t_1, \dots, t_n)})$. Il lemma 3.1.1 assicura che questa famiglia di misure di probabilità soddisfa le condizioni di consistenza di Kolmogorov, dunque il teorema 3.1.1 dimostra l'esistenza del processo $\{X(t)\}_{t \in T}$. □

Il caso che ci interessa di più è quello di un processo gaussiano stazionario (nella proposizione 3.1.2 dimostreremo che stazionarietà in senso lato e in senso stretto coincidono per i processi gaussiani) e a tempi discreti (tipicamente avremo $T = \mathbb{N}$ oppure $T = \mathbb{Z}$). In questa situazione per ogni funzione $\Gamma : \mathbb{Z} \rightarrow \mathbb{R}$ semidefinita positiva ed ogni $m \in \mathbb{R}$ esiste un processo gaussiano $\{X(t)\}_{t \in T}$, con funzione di covarianza Γ e media m .

La proposizione seguente autorizza a parlare genericamente di stazionarietà per i processi gaussiani, senza specificare se essa sia intesa debolmente o in senso stretto.

Proposizione 3.1.2. *Sia $\{X(t)\}_{t \in T}$ un processo gaussiano: esso è debolmente stazionario se e solo se è strettamente stazionario.*

Dimostrazione. Dobbiamo dimostrare che se $\{X(t)\}_{t \in T}$ è debolmente stazionario allora è anche strettamente stazionario. Consideriamo la n -upla di elementi di T (t_1, \dots, t_n) e $h \in \mathbb{R}$, in modo che

$$t_i + h \in T \quad \forall i = 1, \dots, n.$$

Per la stazionarietà in senso lato abbiamo:

$$\text{cov}(X(t_i), X(t_j)) = \text{cov}(X(t_i + h), X(t_j + h)),$$

dunque i vettori aleatori

$$(X(t_1), \dots, X(t_n)) \quad (X(t_1 + h), \dots, X(t_n + h))$$

hanno uguale matrice di covarianza e vettore delle medie. Questo conclude la dimostrazione, poiché questi parametri caratterizzano in modo univoco la distribuzione di un vettore gaussiano. \square

Vediamo ora alcune proprietà generali dei sistemi gaussiani. Ricordiamo che dato uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, l'insieme $L^2(\Omega, \mathcal{F}, \mathbb{P})$ delle variabili aleatorie di quadrato integrabile su di esso è in maniera ovvia uno spazio vettoriale. Inoltre può essere definito su di esso un prodotto scalare

$$\langle X, Y \rangle = \mathbb{E}\{XY\} \quad \text{per } X, \quad Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$$

che rende $L^2(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di Hilbert. Su questo spazio considereremo la topologia indotta dal prodotto scalare definito in precedenza. Dato $W \subset L^2(\Omega, \mathcal{F}, \mathbb{P})$, indicheremo con $\text{span}W$ il sottospazio vettoriale generato da W e con $\overline{\text{span}W}$ la sua chiusura.

Proposizione 3.1.3. Sia $\{X_n\}_{n \in \mathbb{N}}$ una successione di variabili gaussiane definite su uno stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Se esiste una variabile aleatoria Y tale che

$$X_n \xrightarrow{\mathcal{L}} Y$$

allora Y è ancora una variabile gaussiana, con media

$$\lim_n \mathbb{E}\{X_n\}$$

e varianza

$$\lim_n \text{Var}\{X_n\}.$$

Se A è un insieme e $\{X(a) | a \in A\}$ è un sistema gaussiano definito su $(\Omega, \mathcal{F}, \mathbb{P})$, allora anche $\overline{\text{span}}A$ è un sistema gaussiano.

Dimostrazione. Proviamo dapprima l'affermazione riguardante la convergenza in legge. Per il teorema di Lévy sulla convergenza delle funzioni caratteristiche,

$$X_n \xrightarrow{\mathcal{L}} Y$$

se e solo se puntualmente si ha

$$\phi_{X_n} \rightarrow \phi_Y,$$

dove ϕ_Y è la funzione caratteristica della variabile Y . Se poniamo $m_n = \mathbb{E}\{X_n\}$ e $\sigma_n^2 = \text{Var}\{X_n\}$, abbiamo che:

$$\phi_{X_n}(t) = e^{im_n t - \frac{t^2 \sigma_n^2}{2}}.$$

Vogliamo provare che esistono m e σ^2 tali che $m_n \rightarrow_n m$ e $\sigma_n^2 \rightarrow \sigma^2$, perché se questo fosse vero avremmo

$$e^{im_n t - \frac{t^2 \sigma_n^2}{2}} \rightarrow e^{imt - \frac{t^2 \sigma^2}{2}},$$

dunque Y sarebbe una variabile gaussiana con media m e varianza σ^2 .

Imponendo la convergenza della norma delle funzioni caratteristiche per $t = 1$, otteniamo che:

$$\|\phi_{X_n}(1)\| = \|e^{im_n - \frac{\sigma_n^2}{2}}\| = e^{-\frac{\sigma_n^2}{2}} \rightarrow \|\phi_Y(1)\|.$$

Quindi la successione σ_n^2 converge, chiamiamo σ^2 il suo limite. Per il teorema di Lévy si ha allora:

$$e^{im_n} = e^{im_n - \frac{\sigma_n^2}{2}} \cdot e^{\frac{\sigma_n^2}{2}} = \phi_{X_n}(1) \cdot e^{\frac{\sigma_n^2}{2}} \rightarrow \phi_Y(1) e^{\frac{\sigma^2}{2}},$$

dunque anche la successione m_n converge: abbiamo concluso la prima parte della dimostrazione.

La seconda affermazione deriva direttamente dalla prima: è banale infatti verificare che se A è un sistema gaussiano lo è anche $\text{span}A$; di conseguenza anche $\overline{\text{span}A}$ è ancora un sistema gaussiano per la prima parte della dimostrazione. \square

3.2 Filtri lineari

3.2.1 Definizioni

In questo paragrafo descriviamo una classe di processi stocastici a tempi discreti che saranno particolarmente utili nei prossimi capitoli. Alcune condizioni sufficienti affinché un processo gaussiano appartenga a tale classe saranno esaminate nel paragrafo 3.2.3.

Di qui in avanti fisseremo uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e ci riferiremo sempre a processi a valori complessi definiti su di esso. Estenderemo inoltre il significato usuale del termine successione ad una funzione avente come dominio \mathbb{Z} o $\mathbb{Z} \times \mathbb{Z}$.

Definizione 3.2.1. *Sia $\{c_{t,k}\}_{t,k \in \mathbb{Z}}$ una successione di numeri complessi. Un processo $\{X(t)\}_{t \in \mathbb{Z}}$ si dice ottenuto da $\{Z(k)\}_{k \in \mathbb{Z}}$ mediante l'applicazione del filtro lineare $\{c_{t,k}\}_{t,k \in \mathbb{Z}}$ se vale*

$$X(t) = \sum_{k=-\infty}^{\infty} c_{t,k} Z(k), \quad (3.1)$$

dove la convergenza è intesa in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Il filtro $\{c_{t,k}\}$ si dice *invariante* nel tempo se per ogni $k, j, s \in \mathbb{Z}$ è verificata la condizione

$$c_{k+j,k} = c_{s+j,s}$$

In questo caso si ha:

$$X(t-s) = \sum_{k=-\infty}^{\infty} c_{t-s,k} Z(k) = \sum_{k=-\infty}^{\infty} c_{t,k} Z(k-s).$$

Possiamo definire la successione

$$h_j = c_{t+j,t}$$

che chiameremo ancora *filtro*, la (3.1) quindi può essere riscritta come:

$$X(t) = \sum_{k=-\infty}^{\infty} h_k Z(t-k).$$

Il filtro h_j si dice *casuale* se $h_k = 0$ per ogni $k < 0$.

Lemma 3.2.1. *Supponiamo che il processo $\{Z(k)\}_{k \in \mathbb{Z}}$ sia stazionario in senso lato, centrato e con funzione di covarianza Γ_Z . Supponiamo inoltre che la successione $\{h_j\}_{j \in \mathbb{Z}}$ sia assolutamente sommabile, cioè che*

$$\sum_{j=-\infty}^{\infty} \|h_j\| < \infty;$$

allora la serie

$$\sum_{k=-\infty}^{\infty} h_k Z(t-k) \tag{3.2}$$

converge in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ e quasi certamente ad una stessa variabile aleatoria finita.

Dimostrazione. Per il teorema di Beppo Levi abbiamo che:

$$\begin{aligned} \mathbb{E}\left\{ \sum_{k=-\infty}^{\infty} \|h_k\| \|Z(t-k)\| \right\} &= \mathbb{E}\left\{ \lim_{n \rightarrow \infty} \sum_{k=-n}^n \|h_k\| \|Z(t-k)\| \right\} = \\ &= \lim_{n \rightarrow \infty} \sum_{k=-n}^n \|h_k\| \mathbb{E}\{\|Z(t-k)\|\} = \mathbb{E}\{\|Z(t)\|\} \sum_{k=-\infty}^{\infty} \|h_k\|. \end{aligned}$$

Dunque

$$\sum_{k=-\infty}^{\infty} \|h_k\| \|Z(t-k)\|$$

è quasi certamente una variabile aleatoria finita: la serie $\sum_{k=-\infty}^{\infty} h_k Z(t-k)$ converge assolutamente, perciò anche puntualmente ad una variabile aleatoria $X(t)$. Dato che

$$\|X(t)\| \leq \sum_{k=-\infty}^{\infty} \|h_k\| \|Z(t-k)\| = \|Z(t)\| \sum_{k=-\infty}^{\infty} \|h_k\| < \infty$$

anche $X(t)$ è quasi certamente finita. Ora dimostriamo che la serie converge anche in $L^2(\Omega, \mathcal{F}, \mathbb{P})$; ricordando che $\|\Gamma_Z(j)\| \leq \Gamma_Z(0)$ per ogni $j \in \mathbb{Z}$:

$$\mathbb{E}\left\{\left[\sum_{j=m}^n h_j Z(t-j)\right]^2\right\} \leq \sum_{j=m}^n \|h_j\|^2 \Gamma_Z(0) = \Gamma_Z(0) \sum_{m \leq j \leq n} \|h_j\|^2 \rightarrow 0;$$

dunque la serie verifica il criterio di Cauchy. Dal momento che $L^2(\Omega, \mathcal{F}, \mathbb{P})$ è uno spazio di Hilbert, la serie converge ad una variabile aleatoria $X'(t)$. Dimostriamo infine che le due variabili aleatorie coincidono: per il lemma di Fatou

$$\begin{aligned} \mathbb{E}\{(X(t) - X'(t))^2\} &= \mathbb{E}\left\{\liminf_{n \rightarrow \infty} \left\|X(t) - \sum_{k=-n}^n h_k Z(t-k)\right\|^2\right\} \leq \\ &\leq \liminf_{n \rightarrow \infty} \mathbb{E}\left\{\left\|X(t) - \sum_{k=-n}^n h_k Z(t-k)\right\|^2\right\} = 0. \end{aligned}$$

□

Si può descrivere anche la funzione di covarianza di un processo generato dall'applicazione di un filtro lineare, invariante nel tempo e assolutamente sommabile.

Proposizione 3.2.1. *Supponiamo che il processo $\{Z(k)\}_{k \in \mathbb{Z}}$ sia stazionario in senso lato, centrato e con funzione di covarianza Γ_Z . Supponiamo inoltre che la successione $\{h_j\}_{j \in \mathbb{Z}}$ sia assolutamente sommabile, cioè*

$$\sum_{j=-\infty}^{\infty} \|h_j\| < \infty.$$

Sia

$$X(n) = \sum_{j=-\infty}^{\infty} h_j Z(n-j).$$

Il processo $X(n)$ è stazionario in senso lato ed ha funzione di covarianza

$$\Gamma_X(h) = \sum_{j,k=-\infty}^{\infty} h_j \overline{h_k} \Gamma_Z(h-j+k). \quad (3.3)$$

Dimostrazione. Verifichiamo dapprima la stazionarietà del processo $\{X(n)\}_{n \in \mathbb{Z}}$. La serie (3.2) converge in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, quindi possiamo calcolarne il valore atteso:

$$\mathbb{E}\{X(n)\} = \lim_{m \rightarrow \infty} \sum_{j=-m}^m h_j \mathbb{E}\{Z(n-j)\} = 0.$$

Per quanto riguarda la covarianza:

$$\begin{aligned} \mathbb{E}\{X(n+h)\overline{X(n)}\} &= \lim_{m \rightarrow \infty} \mathbb{E}\left\{\left(\sum_{j=-m}^m h_j Z(n+h-j)\right)\left(\sum_{k=-m}^m \overline{h_k Z(n-k)}\right)\right\} = \\ &= \lim_{m \rightarrow \infty} \sum_{j,k=-m}^m h_j \overline{h_k} \Gamma_Z(h-j+k) = \sum_{j,k=-\infty}^{\infty} h_j \overline{h_k} \Gamma_Z(h-j+k). \end{aligned}$$

Prima dell'ultimo passaggio al limite bisognerebbe dimostrare che esso ha senso: questo è molto semplice perché, per l'assoluta sommabilità del filtro

$$\lim_{m \rightarrow \infty} \sum_{j,k=-m}^m \|h_j\| \|h_k\| = \lim_{m \rightarrow \infty} \sum_{j=-m}^m \|h_j\| \lim_{m \rightarrow \infty} \sum_{k=-m}^m \|h_k\| < \infty;$$

quindi la successione $h_j \overline{h_k}$ è assolutamente sommabile; dal momento che $\|\Gamma_Z(n)\| \leq \Gamma_Z(0)$ per ogni $n \in \mathbb{Z}$, anche la successione

$$h_j \overline{h_k} \Gamma_Z(h-j+k)$$

è assolutamente sommabile. Il processo $\{X(n)\}_{n \in \mathbb{Z}}$ quindi è stazionario in senso lato e vale (3.3), come volevamo dimostrare. \square

3.2.2 I processi ARMA(p,q)

In questo e nel prossimo paragrafo vogliamo dare delle condizioni sufficienti affinché un processo $\{X(n)\}_{n \in \mathbb{Z}}$ abbia una rappresentazione come in (3.2). Qui non ipotizzeremo che il processo sia gaussiano: i risultati valgono in generale per processi debolmente stazionari a tempi discreti.

Diamo innanzitutto la definizione di processo ARMA e di rumore bianco.

Definizione 3.2.2. Sia $\{X(n)\}_{n \in \mathbb{Z}}$ un processo stazionario in senso lato definito su $(\Omega, \mathcal{F}, \mathbb{P})$, con funzione di covarianza Γ . Esso si dice rumore bianco se

$$\Gamma(j) = 0 \quad \text{se } j \neq 0.$$

Definizione 3.2.3. Siano $a_1, \dots, a_m \in \mathbb{C}$ e $b_1, \dots, b_k \in \mathbb{C}$. Un processo centrato $\{X(n)\}_{n \in \mathbb{Z}}$ definito su $(\Omega, \mathcal{F}, \mathbb{P})$ si dice ARMA(p, q) se esso è stazionario in senso lato ed è soluzione dell'equazione alle differenze:

$$X(n) - a_1 X(n-1) - \dots - a_m X(n-m) = Z(n) + b_1 Z(n-1) + \dots + b_k Z(n-k), \quad (3.4)$$

dove $\{Z(n)\}_{n \in \mathbb{Z}}$ è un rumore bianco di varianza σ^2 definito su $(\Omega, \mathcal{F}, \mathbb{P})$. Un processo $\{X(n)\}_{n \in \mathbb{Z}}$ di media $\mu(n)$ si dice ARMA se $\{X(n) - \mu(n)\}_{n \in \mathbb{Z}}$ lo è.

Vogliamo dimostrare che, in determinate condizioni, i processi ARMA esistono e determinare una loro rappresentazione. Poniamo:

$$B(X(n)) = X(n-1) \quad \text{e} \quad B(Z(n)) = Z(n-1).$$

B si dice *operatore ritardo*: esso non è un operatore vettoriale, ma solo un formalismo introdotto per evidenziare le analogie fra filtri lineari e serie di potenze.

Con l'introduzione di B possiamo riscrivere la nostra equazione alle differenze come:

$$X(n) - a_1 B(X(n)) - \dots - a_m B^m(X(n)) = Z(n) + b_1 B(Z(n)) + \dots + b_k B^k(Z(n)).$$

Definendo i polinomi a coefficienti complessi

$$a[x] = 1 - a_1 x - \dots - a_m x^m \quad b[x] = 1 + b_1 x + \dots + b_k x^k,$$

l'equazione può essere riscritta in forma compatta come

$$a(B)X(n) = b(B)Z(n).$$

Vediamo alcune proprietà dell'operatore ritardo B . Supponiamo che

$$\phi = \{\phi_j\}_{j \in \mathbb{Z}}$$

sia una successione di numeri complessi assolutamente sommabile. Il lemma 3.2.1 garantisce che ha senso anche l'operatore

$$\phi(B) = \sum_{i=-\infty}^{\infty} \phi_j B^j,$$

se applicato ad un processo stazionario in senso lato.

Se $\alpha = \{\alpha_j\}_{j \in \mathbb{Z}}$ e $\beta = \{\beta_j\}_{j \in \mathbb{Z}}$ sono due successioni di numeri complessi assolutamente sommabili, allora ha senso ed è assolutamente sommabile anche la successione

$$\psi_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{k=-\infty}^{\infty} \alpha_{j-k} \beta_k,$$

prodotto di Cauchy delle successioni α_j e β_j , e si ha:

$$\alpha(B) \circ \beta(B)X(t) = \beta(B) \circ \alpha(B)X(t) = \psi(B)X(t).$$

Proposizione 3.2.2. *Se il polinomio $a[x]$ non ha radici complesse di modulo unitario, l'unico processo stazionario in senso lato $\{X(n)\}_{n \in \mathbb{N}}$ definito su $(\Omega, \mathcal{F}, \mathbb{P})$ che verifica l'equazione alle differenze 3.4 è dato da:*

$$X(n) = \sum_{j=-\infty}^{\infty} \psi_j Z(n-j),$$

dove i ψ_j sono i coefficienti dello sviluppo in serie di Laurent della funzione $b[x]a[x]^{-1}$.

Dimostrazione. Per ipotesi $a[x]$ non ha radici complesse di modulo unitario: esiste quindi $R > 1$ tale che la funzione $b[x]a[x]^{-1}$ è analitica sull'anello

$$A(0, R^{-1}, R) = \{x \in \mathbb{C} | R^{-1} \leq \|x\| \leq R\}.$$

Il teorema di Laurent garantisce che per ogni $x \in A(0, R^{-1}, R)$, $b[x]a^{-1}[x]$ ha una rappresentazione in serie di Laurent

$$\sum_{j=-\infty}^{j=\infty} \psi_j x^j.$$

e che tale serie converga assolutamente. Siccome la circonferenza unitaria è un sottoinsieme di $A(0, R^{-1}, R)$, la convergenza assoluta della serie su x con $\|x\| = 1$ implica che

$$\sum_{j=-\infty}^{j=\infty} \|\psi_j\| < \infty.$$

Per la proposizione 3.2.1 $X(n)$ è un processo stazionario in senso lato. Per quanto provato prima a proposito della corrispondenza fra composizione e prodotto di Cauchy delle successioni, abbiamo:

$$a(B) \circ \psi(B)Z(t) = b(B)Z(t),$$

dunque $X(n)$ è effettivamente una soluzione stazionaria dell'equazione alle differenze.

Per provare che essa è anche unica, supponiamo che $\{X'(n)\}_{n \in \mathbb{Z}}$ sia un'altra soluzione stazionaria. Per l'ipotesi su $a[x]$ e il teorema di Laurent, $a[x]^{-1}$ può essere sviluppata in serie di Laurent in un anello contenente la sfera unitaria. Siano ϵ_j i coefficienti dello sviluppo. Può essere definito l'operatore $\epsilon(B)$ ed applicandolo ad ambo i membri della nostra equazione otteniamo:

$$X'(n) = \epsilon(B) \circ a(B)X'(n) = \epsilon(B) \circ b(B)Z(n) = \psi(B)Z(n) = X(n).$$

□

3.2.3 I processi gaussiani puramente non deterministici

Con un approccio mutuato dalla teoria della predizione si riescono ad ottenere delle condizioni sufficienti affinché un processo gaussiano possa essere rappresentato come in (3.2). Introduciamo dapprima la terminologia necessaria.

Supponiamo che $\{Y(m)\}_{m \in \mathbb{Z}}$ sia un processo stocastico a valori reali, stazionario in senso lato e definito su $(\Omega, \mathcal{F}, \mathbb{P})$. Consideriamo sull'insieme $L^2(\Omega, \mathcal{F}, \mathbb{P})$ la topologia indotta dal prodotto scalare

$$\langle X, Y \rangle = \mathbb{E}\{X \cdot Y\} \quad \forall X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$$

. Poniamo

$$H_n(X) = \overline{\text{span}}\{X_k | k \leq n\}.$$

Per ogni $n \in \mathbb{Z}$ vale l'inclusione:

$$H_n(X) \subseteq H_{n+1}(X).$$

Se il processo verifica $H_n(X) = H_{n+1}(X)$ per ogni n esso si dice *deterministico*, in caso contrario si dice *non deterministico*. Se il processo non è formato da variabili quasi certamente costanti, il caso:

$$\bigcap_{n \in \mathbb{Z}} H_n(X) \subseteq \text{span}\{1\}, \quad (3.5)$$

è un caso particolare di non determinismo: se (3.5) è soddisfatta il processo $\{X(n)\}_{n \in \mathbb{N}}$ si dice *puramente non deterministico*.

Spieghiamo brevemente il senso di queste definizioni: dato il processo $\{X(n)\}_{n \in \mathbb{Z}}$, la teoria della predizione si occupa della previsione della variabile $X(t+h)$, note le variabili $\{X(n) | n \leq t\}$. Supponiamo di voler scegliere la nostra variabile predittiva $\hat{X}(t+h)$ in $H_t(X)$: se per ogni n abbiamo che $H_n(X) = H_{n+1}(X)$, è possibile scegliere una previsione esatta, per cui cioè $\hat{X}(t+h) = X(t+h)$. In questo caso dunque il processo si dice deterministico, perché è possibile determinare il futuro $X(t+h)$ a partire dal passato $\{X(n) | n \leq t\}$. Nel caso contrario, quella data non sarà una previsione esatta, perché in generale $\hat{X}(t+s) \neq X(t+s)$ e, per questo, il processo si dice non deterministico.

D'ora in avanti supporremo che $\{X(n)\}_{n \in \mathbb{Z}}$ sia gaussiano, centrato, stazionario e puramente non deterministico. Dividiamo il paragrafo in due parti: nella prima determiniamo il processo $\{Z(n)\}_{n \in \mathbb{Z}}$, che, a meno di piccole modifiche, servirà per ottenere la rappresentazione

$$X(n) = \sum_{j=-\infty}^{\infty} h_j Z(n-j);$$

nella seconda dimostreremo l'esistenza della successione $\{h_j\}_{j \in \mathbb{Z}}$.

Prima della costruzione del processo $\{Z(n)\}_{n \in \mathbb{Z}}$ dimostriamo due lemmi.

Lemma 3.2.2. *Se $H_n(X)^\perp$ è il complemento ortogonale di $H_n(X)$ in $H_{n+1}(X)$, allora:*

$$\dim H_n(X)^\perp \leq 1,$$

inoltre $\dim H_n(X)^\perp$ non dipende da n .

Dimostrazione. Sappiamo che

$$H_{n+1}(X) = \overline{\text{span}}\{H_n(X), X_{n+1}\} = \text{span}\{H_n(X), X_{n+1}\},$$

quindi, indicando con π la proiezione di $H_{n+1}(X)$ sul sottospazio chiuso $H_n(X)$, il vettore

$$X(n+1) - \pi(X(n+1))$$

è un generatore di $H_n(X)^\perp$: la prima affermazione è dimostrata.

Supponiamo poi che per un $m \in \mathbb{N}$ valga l'uguaglianza

$$\pi(X(m+1)) = X(m+1),$$

dunque che $\dim H_m(X)^\perp = 0$. Questo è vero se e solo se esiste una successione $\{v_n\}_{n \in \mathbb{N}}$ di elementi di

$$\text{span}\{X(k) | k \leq m\}$$

tale che per $n \rightarrow \infty$

$$\|v_n - X(m+1)\|^2 \rightarrow 0.$$

Siccome $v_n \in \text{span}\{X(k) | k \leq m\}$, esistono d_n interi $k_1, \dots, k_{d_n} \leq m$ e $a_1, \dots, a_{d_n} \in \mathbb{R}$ tali che:

$$v_n = a_1 X(k_1) + \dots + a_{d_n} X(k_{d_n}).$$

Sia $f \in \mathbb{Z}$; indicando con B l'operatore ritardo sul processo $\{X(n)\}_{n \in \mathbb{Z}}$, consideriamo i vettori

$$B^{-f}(v_n) = a_1 B^{-f}(X(k_1)) + \dots + a_{d_n} B^{-f}(X(k_{d_n})).$$

Per la stazionarietà del processo

$$\|B^{-f}(v_n) - X(m+f+1)\|^2 \rightarrow 0,$$

quindi $X(m+f+1) \in H_{m+f}(X)$ e $\dim H_{m+f}(X)^\perp = 0$. □

Lemma 3.2.3. *Sia (V, \langle, \rangle) uno spazio di Hilbert e $\{W_n\}_{n \in \mathbb{N}}$ una successione di sottospazi vettoriali chiusi di V , tale che $W_{n+1} \subset W_n$. Indichiamo con π_n la proiezione di V sul sottospazio W_n e con π quella sul sottospazio $W = \bigcap_{n \in \mathbb{Z}} W_n$, allora per ogni $v \in V$*

$$\lim_{n \rightarrow \infty} \pi_n(v) = \pi(v).$$

Dimostrazione. Dimostriamo dapprima che la successione $\{\pi_n(v)\}_{n \in \mathbb{N}}$ converge. Siccome (V, \langle, \rangle) è uno spazio di Hilbert, condizione necessaria e sufficiente affinché questo accada è che la successione sia di Cauchy. Supponiamo per assurdo che ciò non sia vero e, fissato $\epsilon > 0$, definiamo ricorsivamente una successione di naturali $\{k_n\}_{n \in \mathbb{N}}$: scegliamo k_0 arbitrariamente, poi, fissato k_n , scegliamo k_{n+1} nell'insieme

$$\{z \in \mathbb{N} \mid z > k_n \text{ e } \|\pi_z(v) - \pi_{k_n}(v)\| \geq \epsilon\}, \quad (3.6)$$

che non è vuoto perché abbiamo supposto che la successione $\{\pi_n(v)\}_{z \in \mathbb{N}}$ non sia di Cauchy.

Sia V_n il complemento ortogonale di W_{k_n} in W_{k_0} . Per (3.6), per ogni n , si ha

$$\|v\| \geq \|\pi_{V_n}(v)\| = \sum_{i=0}^{n-1} \|\pi_{k_i}(v) - \pi_{k_{i+1}}(v)\| \geq n\epsilon,$$

il che è assurdo, dal momento che $\|v\| < \infty$.

Sia quindi w il limite della successione $\{\pi_n(v)\}_{n \in \mathbb{N}}$, dimostriamo che $w = \pi(v)$. $w \in W_n$ per ogni n , perché limite di una successione definitivamente contenuta in W_n , dunque $w \in \bigcap_{n \in \mathbb{N}} W_n = W$. Inoltre

$$w - v \in W^\perp$$

perché, per ogni $x \in W$

$$\langle w - v, x \rangle = \lim_{n \rightarrow \infty} \langle \pi_n(v) - v, x \rangle = 0,$$

dal momento che $\pi_n(v) - v \in W_n^\perp$ e $x \in W_n$. La decomposizione

$$v = v_W + v_{W^\perp} \text{ con } v_W \in W \text{ e } v_{W^\perp} \in W^\perp$$

è unica, dunque deve essere $w = \pi(v)$. □

Supponiamo che il processo $\{X(n)\}_{n \in \mathbb{Z}}$ sia puramente non deterministico e che $\dim H_n(X)^\perp = 1$ per ogni n . Scegliamo una successione $\{Z(n)\}_{n \in \mathbb{Z}}$ di elementi di $H(X)$ tale che, per ogni n , si abbia

$$\|Z(n)\| = 1 \quad \text{e} \quad Z(n) \in H_n^\perp.$$

Vogliamo dimostrare che $\{Z(n)\}_{n \in \mathbb{Z}}$ è un sistema ortonormale e completo di $H(X)$. Per la scelta della successione è evidente che, per $m, n \in \mathbb{Z}$

$$\langle Z(n), Z(m) \rangle = \delta_{n,m}$$

dove la funzione $\delta_{n,m}$ è la *delta* di Kronecker. Si tratta quindi di dimostrare che

$$\overline{\text{span}}\{Z(n)|n \in \mathbb{Z}\} = H(X).$$

$Z(n) \in H(X)$ per ogni n , dunque un'inclusione è immediata. Per provare l'inclusione opposta proviamo che, per ogni m

$$X(m) \in \overline{\text{span}}\{Z(n)|n \in \mathbb{Z}\}.$$

Sia $k < m$ e indichiamo con π_k la proiezione di $H_m(X)$ sul sottospazio chiuso $H_k(X)$: possiamo rappresentare $X(m)$ come

$$X(m) = \pi_k(X(m)) + (X(m) - \pi_k(X(m))),$$

dove $(X(m) - \pi_k(X(m))) \in \text{span}\{Z(k), \dots, Z(m-1)\}$.

I sottospazi vettoriali chiusi $\{H_k(X)\}_{k \leq m}$ sono una successione decrescente, con $H_{-\infty}(X) = \bigcap_{k \leq n} H_n(X) \subset \text{span}\{1\}$; le variabili in $H(X)$ sono centrate, dunque nel nostro caso

$$H_{-\infty}(X) = \{0\}$$

Per quanto dimostrato nel lemma 3.2.3, facendo tendere k a $-\infty$

$$\pi_k(X(m)) \xrightarrow{L^2} \pi_{H_{-\infty}(X)}(X(m)) = 0,$$

dunque

$$X(m) - \pi_k(X(m)) \xrightarrow{L^2} X(m).$$

Poiché per ogni k

$$X(m) - \pi_k(X(m)) \in \overline{\text{span}}\{Z(n)|n \in \mathbb{Z}\}$$

e $\overline{\text{span}}\{Z(n)|n \in \mathbb{Z}\}$ è chiuso, anche $X(m) \in \overline{\text{span}}\{Z(n)|n \in \mathbb{Z}\}$.

Il lemma che segue sarà utile per dimostrare l'esistenza del filtro invariante nel tempo $\{h_j\}_{j \in \mathbb{Z}}$.

Lemma 3.2.4. *Siano $h, h' \in \mathbb{N}$, se indichiamo con π_n la proiezione di $H(X)$ sul sottospazio vettoriale chiuso $H_n(X)$, la quantità*

$$\langle X(n+h) - \pi_n(X(n+h)), X(n+h') - \pi_n(X(n+h')) \rangle$$

non dipende da n .

Dimostrazione. Dimostrare la tesi equivale a dimostrare che, per ogni n e $f \in \mathbb{Z}$,

$$\begin{aligned} & \langle X(n+h) - \pi_n(X(n+h)), X(n+h') - \pi_n(X(n+h')) \rangle = \\ & = \langle X(n+f+h) - \pi_{n+f}(X(n+f+h)), X(n+f+h') - \pi_{n+f}(X(n+f+h')) \rangle. \end{aligned}$$

Siccome $\pi_n(X(n+h)) \in H_n(X)$, esistono due successioni $\{v_m\}_{m \in \mathbb{N}}$ e $\{w_m\}_{m \in \mathbb{N}}$ di vettori in $\text{span}\{X(k), k \leq n\}$ tali che

$$v_m \xrightarrow{L^2} \pi_n(X(n+h)) \quad \text{e} \quad w_m \xrightarrow{L^2} \pi_n(X(n+h')).$$

In maniera del tutto analoga a quanto fatto nella dimostrazione del lemma 3.2.2, definiamo le successioni

$$\{B^{-f}(v_n)\}_{n \in \mathbb{N}} \quad \{B^{-f}(w_n)\}_{n \in \mathbb{N}}.$$

Per la stazionarietà del processo $\{X(n)\}_{n \in \mathbb{Z}}$

$$B^{-f}(v_n) \xrightarrow{L^2} \pi_{n+f}(X(n+f+h)) \quad \text{e} \quad B^{-f}(w_n) \xrightarrow{L^2} \pi_{n+f}(X(n+f+h')).$$

Ancora, per la stazionarietà del processo $\{X(n)\}_{n \in \mathbb{N}}$:

$$\langle X(n+h) - v_m, X(n+h') - w_m \rangle = \langle X(n+f+h) - B^{-f}(v_m), X(n+f+h') - B^{-f}(w_m) \rangle.$$

Passando al limite su m , per la continuità del prodotto scalare si ottiene:

$$\begin{aligned} & \langle X(n+h) - \pi_n(X(n+h)), X(n+h') - \pi_n(X(n+h')) \rangle = \\ & = \langle X(n+f+h) - \pi_{n+f}(X(n+f+h)), X(n+f+h') - \pi_{n+f}(X(n+f+h')) \rangle. \end{aligned}$$

□

Ora dimostriamo l'esistenza del filtro invariante nel tempo.

Proposizione 3.2.3. *Supponiamo che $\{X(n)\}_{n \in \mathbb{Z}}$ sia puramente non deterministico e che lo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ sia non atomico. Esistono un processo $\{Z(t)\}_{t \in \mathbb{Z}}$ definito su $(\Omega, \mathcal{F}, \mathbb{P})$ ed un filtro $\{h_j\}_{j \in \mathbb{N}}$ tali che:*

$$X(n) = \sum_{j=-\infty}^{\infty} h_j Z(n-j).$$

Inoltre il filtro è casuale e il processo $Z(t)$ è formato da variabili gaussiane standard indipendenti.

Dimostrazione. Per il lemma 3.2.2 sappiamo che $\dim H_n(X)^\perp \leq 1$ e che essa è indipendente da n .

Se $\dim H_n(X)^\perp = 0$ avremmo che $X(n) \in H_m(X)$ per ogni $m \in \mathbb{Z}$, ma allora $X(n) \in H_{-\infty}(X) \subset \text{span}\{1\}$ e, visto che si tratta di una variabile centrata,

$$X(n) = 0 \quad \forall n \in \mathbb{Z}.$$

In questo caso abbiamo terminato: definita una successione di variabili gaussiane standard ed indipendenti (questo è sempre possibile su uno spazio di probabilità non atomico), basta porre $h_j = 0$ per ogni $j \in \mathbb{Z}$.

Supponiamo ora che $\dim H_n(X)^\perp = 1$; consideriamo un sistema $\{Z(n)\}_{n \in \mathbb{Z}}$ ortonormale e completo di $H(X)$ tale che $Z(n) \in H_n(X)^\perp$ (nella prima parte del paragrafo è stato dimostrato che un simile sistema ortonormale completo esiste): per la proposizione 3.1.3 gli $H_n(X)$ sono sistemi gaussiani, dunque il processo $\{Z(n)\}_{n \in \mathbb{Z}}$ è formato da variabili gaussiane standard (perché centrate e tali che $\|Z(n)\| = 1$) ed indipendenti (perché non correlate).

L'insieme $\{Z(j) | j \leq n\}$ costituisce un sistema ortonormale completo per il sottospazio chiuso $H_n(X)$. Dunque per ogni $n \in \mathbb{Z}$ esiste una successione $\{a_{n,j}\}_{j \in \mathbb{Z}}$ tale che

$$\sum_{j \in \mathbb{Z}} |a_{n,j}| < \infty \quad \text{e} \quad \sum_{j=-\infty}^n a_{n,j} Z(j) = X(n);$$

possiamo porre $a_{n,j} = 0$ se $j > n$ e, a meno di cambiare di segno i vettori del sistema ortonormale $\{Z(n)\}_{n \in \mathbb{Z}}$, supporre che $a_{n,n} > 0$ per ogni $n \in \mathbb{Z}$.

Vogliamo dimostrare che il filtro $a_{n,j}$ è invariante nel tempo, ossia che per ogni coppia di interi m, n e per ogni $j > 0$

$$a_{m+j,m} = a_{n+j,n}.$$

Se questo fosse vero potremmo definire una successione h_j , come fatto nel paragrafo 3.2.1, tale che

$$X(n) = \sum_{j=-\infty}^n h_j Z(n-j) \quad \forall n \in \mathbb{Z}.$$

Indichiamo con π_n la proiezione di $H(X)$ sul sottospazio chiuso $H_n(X)$. Dal momento che in $(\Omega, \mathcal{F}, \mathbb{P})$ la proiezione coincide con la speranza condizionale

$$\pi_n : Y \mapsto \mathbb{E}\{Y | X(k), k \leq n\},$$

la costruzione del sistema ortonormale $\{Z(n)\}_{n \in \mathbb{Z}}$ implica che se $q > p$ le variabili $Z(q)$ e $X(p)$ sono indipendenti, mentre se $q < p$ la variabile $Z(q)$ è misurabile rispetto alla tribù generata da $H_p(X)$.

Per la continuità della proiezione ortogonale in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ e per le note proprietà della speranza condizionale, se $k > n$:

$$\begin{aligned} \pi_n(X(k)) &= \mathbb{E}\left\{ \sum_{j=-\infty}^k a_{k,j} Z(j) \middle| X(m), m \leq n \right\} = \sum_{j=-\infty}^k \mathbb{E}\{a_{k,j} Z(j) \mid X(m), m \leq n\} = \\ &= \sum_{j=-\infty}^n a_{k,j} Z(j). \end{aligned}$$

Dato $h > 0$, consideriamo le variabili aleatorie

$$X(n) - \pi_{n-1}(X(n)) = \sum_{j=-\infty}^n a_{n,j} Z(j) - \sum_{j=-\infty}^{n-1} a_{n,j} Z(j) = a_{n,n} Z(n)$$

$$X(n+h) - \pi_n(X(n+h)) = \sum_{j=-\infty}^{n+h} a_{n+h,j} Z(j) - \sum_{j=-\infty}^n a_{n+h,j} Z(j) = \sum_{j=n}^{n+h} a_{n+h,j} Z(j).$$

Poiché $\dim H_{n-1}(X) = 1$, deve essere $a_{n,n} \neq 0$. Per il lemma 3.2.4 le quantità

$$a_{n,n}^2 = \langle X(n) - \pi_{n-1}(X(n)), X(n) - \pi_{n-1}(X(n)) \rangle$$

$$\sum_{j=n+1}^{n+h'} a_{n+h,j} a_{n+h',j} = \langle X(n+h) - \pi_n(X(n+h)), X(n+h') - \pi_n(X(n+h')) \rangle \quad \text{con } h > h' > 0$$

non dipendono da n .

Verifichiamo per induzione su $j \in \mathbb{N}$ che per ogni n, m si ha

$$a_{n+j,n} = a_{m+j,m}$$

Se $j = 0$ la tesi deriva direttamente dal fatto che $a_{n,n}^2$ non dipende da n e che $a_{n,n} > 0$. Supponiamo dunque che l'ipotesi valga per $j \geq 0$ e verifichiamola per $j + 1$. Posti $h = k + 2$ e $h' = 1$, la quantità

$$\sum_{j=n+1}^{n+h'} a_{n+h,j} a_{n+h',j}$$

non dipende da n ed è un'equazione lineare in $a_{n+k+2,n+1}$, che ha un'unica soluzione, perché $a_{n,n}$, coefficiente dell'incognita, è diverso da 0. \square

3.3 I processi gaussiani di Markov

Lo scopo di questa sezione è caratterizzare i processi gaussiani di Markov in base alla loro funzione di covarianza, supposto che essa sia definita positiva. Ricordiamo che cosa sono vettori e processi di Markov.

Definizione 3.3.1. *Sia (X_1, \dots, X_m) un vettore aleatorio a valori in \mathbb{R}^m , esso si dice un vettore di Markov se per ogni $1 < s \leq m$ e per ogni $A \subset \mathbb{R}$ misurabile*

$$\mathbb{E}\{I_A \circ X_s | X_1, \dots, X_{s-1}\} = \mathbb{E}\{I_A \circ X_s | X_{s-1}\}.$$

La definizione implica che una versione della legge di X_s condizionale a (X_1, \dots, X_{s-1}) e una versione della legge di X_s condizionale a X_{s-1} coincidano \mathbb{P} -quasi certamente. Supponiamo che $X_i \in L^1((\Omega, \mathcal{F}, \mathbb{P}))$ per ogni $i = 1, \dots, m$: il vettore aleatorio

$$(X_1, \dots, X_m)$$

è di Markov se e solo se lo è il vettore

$$(X_1 - \mathbb{E}\{X_1\}, \dots, X_m - \mathbb{E}\{X_m\})$$

Senza perdere di generalità, dunque, possiamo considerare solo vettori aleatori centrati.

Definizione 3.3.2. *Sia $\{X(t)\}_{t \in T}$ un processo stocastico definito su $(\Omega, \mathcal{F}, \mathbb{P})$. Esso si dice di Markov se per ogni $m \in \mathbb{N}$ e per ogni $t_1 < \dots < t_m \in T$ il vettore aleatorio*

$$(X(t_1), \dots, X(t_m))$$

è un vettore di Markov.

Per quanto detto a proposito dei vettori di Markov, se consideriamo processi a variabili integrabili possiamo supporre, senza perdere di generalità, che essi siano centrati.

3.3.1 Vettori gaussiani di Markov

In questo paragrafo studieremo i vettori gaussiani di Markov, che saranno utili, visto il legame esistente fra le definizioni 3.3.1 e 3.3.2, a descrivere i processi gaussiani di Markov. Per quanto detto nel paragrafo precedente, ci limiteremo a considerare il caso di vettori aleatori centrati.

Proposizione 3.3.1. *Supponiamo che (X_1, \dots, X_m) sia un vettore gaussiano centrato, con matrice di covarianza*

$$Q = \{q_{i,j}\}_{i,j=1,\dots,m}$$

definita positiva. Esistono $a_1, \dots, a_{m-1} \in \mathbb{R}$ tali che

$$\mathbb{E}\{X_m | X_1, \dots, X_{m-1}\} = \sum_{i=1}^{m-1} a_i X_i.$$

Una versione della legge di X_m condizionale a $X_1 = x_1, \dots, X_{m-1} = x_{m-1}$ è ancora gaussiana, con media $\sum_{i=1}^{m-1} a_i x_i$ e varianza

$$(1, -a_1, \dots, -a_{m-1})Q(1, -a_1, \dots, -a_{m-1})^t.$$

Dimostrazione. Consideriamo il seguente sistema, nelle incognite x_1, \dots, x_{m-1} :

$$\left\{ \begin{array}{l} a_1 q_{1,1} + a_2 q_{1,2} + \dots + a_{m-1} q_{1,m-1} = q_{1,m} \\ a_1 q_{2,1} + a_2 q_{2,2} + \dots + a_{m-1} q_{2,m-1} = q_{2,m} \\ \dots \\ a_1 q_{m-1,1} + a_2 q_{m-1,2} + \dots + a_{m-1} q_{m-1,m-1} = q_{m-1,m}. \end{array} \right. \quad (3.7)$$

Esso ha un'unica soluzione per il teorema di Rouché-Capelli, perché la matrice Q ha rango massimo, essendo definita positiva. Sia (a_1, \dots, a_{m-1}) la soluzione del sistema e Y la variabile aleatoria

$$Y = X_m - \sum_{i=1}^{m-1} a_i X_i$$

Il vettore aleatorio (Y, X_1, \dots, X_{m-1}) è ancora un vettore gaussiano e Y non è correlata con X_1, \dots, X_{m-1} , perché per la definizione di (a_1, \dots, a_{m-1}) :

$$\langle Y, X_j \rangle = q_{j,m} - \sum_{i=1}^{m-1} a_i q_{j,i} = 0 \quad \forall j = 1, \dots, m-1$$

Siccome (Y, X_1, \dots, X_{m-1}) è un vettore gaussiano, Y è indipendente da X_1, \dots, X_{m-1} . Dunque

$$0 = \mathbb{E}\{Y\} = \mathbb{E}\{Y | X_1, \dots, X_{m-1}\} = \mathbb{E}\{X_m | X_1, \dots, X_{m-1}\} - \sum_{i=1}^{m-1} a_i X_i,$$

da cui

$$\mathbb{E}\{X_m|X_1, \dots, X_{m-1}\} = \sum_{i=1}^{m-1} a_i X_i.$$

La prima affermazione è stata dimostrata. Per le proprietà della speranza condizionale e per l'indipendenza fra Y e X_1, \dots, X_{m-1} , dato $A \subset \mathbb{R}$ misurabile

$$\mathbb{E}\{I_A \circ X_m | X_1, \dots, X_{m-1}\} = \mathbb{E}\{I_A \circ (Y + \sum_{i=1}^{m-1} a_i X_i) | X_1, \dots, X_{m-1}\} = \mathbb{E}\{I_A \circ (Y + \sum_{i=1}^{m-1} a_i x_i)\},$$

dunque una versione della distribuzione di X_m condizionale a $X_1 = x_1, \dots, X_{m-1} = x_{m-1}$ è ancora gaussiana, con media $\sum_{i=1}^{m-1} a_i x_i$ e varianza uguale a quella di Y , ossia

$$(1, -a_1, \dots, -a_{m-1})Q(1, -a_1, \dots, -a_{m-1})^t.$$

□

Nel caso particolare di $m = 2$ si ottiene

$$(1, -a_1)Q(1, -a_1)^t = \frac{q_{1,2}}{q_{1,1}}.$$

Diamo una prima caratterizzazione dei vettori aleatori gaussiani, basata sulla speranza condizionale.

Lemma 3.3.1. *Sia (X_1, \dots, X_m) un vettore aleatorio gaussiano definito su $(\Omega, \mathcal{F}, \mathbb{P})$, con matrice di covarianza non singolare. (X_1, \dots, X_m) è un vettore di Markov se e solo se per ogni $1 < k \leq m$:*

$$\mathbb{E}\{X_k | X_1, \dots, X_{k-1}\} = \mathbb{E}\{X_k | X_{k-1}\}.$$

Dimostrazione. Supponiamo che il vettore (X_1, \dots, X_m) sia di Markov. Indichiamo con $\mathbb{P}(\circ | X_1 = x_1, \dots, X_{k-1} = x_{k-1})$ una versione della legge di X_k condizionale a $X_1 = x_1, \dots, X_{k-1} = x_{k-1}$, che coincide con una versione di $\mathbb{P}(\circ | X_{k-1} = x_{k-1})$, legge di X_k condizionale a X_{k-1} . Allora

$$\begin{aligned} \mathbb{E}\{X_k | X_1, \dots, X_{k-1}\} &= \int_{\mathbb{R}} x \mathbb{P}(dx | X_1(\omega), \dots, X_{k-1}(\omega)) = \\ &= \int_{\mathbb{R}} x \mathbb{P}(dx | X_{k-1}(\omega)) = \mathbb{E}\{X_k | X_{k-1}\}. \end{aligned}$$

Verifichiamo che si tratti anche di una condizione sufficiente. Per il lemma 3.3.1 esistono a_1, \dots, a_{k-1} e b_{k-1} tali che

$$\mathbb{E}\{X_k | X_1, \dots, X_{k-1}\} = \sum_{i=1}^{k-1} a_i X_i$$

$$\mathbb{E}\{X_k | X_{k-1}\} = b_{k-1} X_{k-1}.$$

Dunque, per ipotesi, quasi certamente rispetto a \mathbb{P}

$$(a_{k-1} - b_{k-1})X_{k-1} + \sum_{i=1}^{k-2} a_i X_i = 0.$$

La matrice di covarianza del vettore (X_1, \dots, X_k) è definita positiva, perché anche quella del vettore (X_1, \dots, X_m) lo era: questo implica che la distribuzione di (X_1, \dots, X_k) non possa essere concentrata su alcun iperpiano di \mathbb{R}^k , quindi si deve avere

$$a_{k-1} = b_{k-1} \quad \text{e} \quad a_i = 0 \quad \forall i = 1, \dots, k-2.$$

Dunque, per la proposizione 3.3.1 una versione della legge X_k condizionale a $X_1 = x_1, \dots, X_{k-1} = x_{k-1}$ è gaussiana, con media $a_{k-1}x_{k-1}$ e varianza σ^2 e, d'altra parte, coincide con una versione della legge di X_k condizionale a $X_{k-1} = x_{k-1}$; per ogni $A \subset \mathbb{R}$ misurabile quindi

$$\mathbb{E}\{I_A \circ X_k | X_1, \dots, X_{k-1}\} = \mathbb{E}\{I_A \circ X_k | X_{k-1}\}.$$

□

A questo punto possiamo dare una caratterizzazione dei vettori gaussiani di Markov basata sulla loro funzione di covarianza. Introduciamo dapprima il coefficiente di correlazione secondo Pearson: date due variabili aleatorie $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$,

$$\rho = \frac{\text{cov}\{X, Y\}}{\sqrt{\text{Var}\{X\}\text{Var}\{Y\}}}.$$

Assegnato un vettore aleatorio (X_1, \dots, X_n) con componenti di quadrato integrabile, porremo

$$\rho_{i,j} = \frac{\text{cov}(X_i, X_j)}{\sqrt{\text{Var}\{X_i\}\text{Var}\{X_j\}}}.$$

Proposizione 3.3.2. Sia (X_1, \dots, X_m) un vettore gaussiano definito su $(\Omega, \mathcal{F}, \mathbb{P})$ con matrice di covarianza $Q = \{q_{i,j}\}_{i,j=1,\dots,m}$ non singolare. Esso è di Markov se e solo se per ogni terna di indici $1 \leq j \leq v \leq k \leq n$ vale

$$\rho(j, k) = \rho(j, v)\rho(v, k).$$

Dimostrazione. Ripercorriamo la dimostrazione della proposizione 3.3.1: esiste un'unica scelta di coefficienti $a_1, \dots, a_{k-1} \in \mathbb{R}$ tale che

$$\mathbb{E}\{X_k | X_1, \dots, X_{k-1}\} = \sum_{i=1}^{k-1} a_i X_i,$$

e l'uguaglianza vale se e solo se la variabile aleatoria

$$G = X_k - \sum_{i=1}^{k-1} a_i X_i$$

è indipendente da X_1, \dots, X_{k-1} . Per la stessa proposizione inoltre

$$\mathbb{E}\{X_k | X_1, \dots, X_{k-1}\} = \frac{q_{k-1,k}}{q_{k,k}} X_{k-1} = \frac{\sqrt{\text{Var}\{X_k\}}}{\sqrt{\text{Var}\{X_{k-1}\}}} \rho_{k-1,k} X_{k-1}.$$

Quindi il vettore (X_1, \dots, X_m) è di Markov se e solo se, per il lemma 3.3.1,

$$\mathbb{E}\{X_k | X_1, \dots, X_{k-1}\} = \frac{\sqrt{\text{Var}\{X_k\}}}{\sqrt{\text{Var}\{X_{k-1}\}}} \rho_{k-1,k} X_{k-1}.$$

Condizione necessaria e sufficiente affinché questo accada è che

$$X_k - \frac{\sqrt{\text{Var}\{X_k\}}}{\sqrt{\text{Var}\{X_{k-1}\}}} \rho_{k-1,k} X_{k-1}$$

non sia correlata con X_1, \dots, X_{k-1} , che è equivalente alla condizione

$$\rho_{i,k} = \rho_{i,k-1} \rho_{k-1,k}. \quad (3.8)$$

Presi $1 \leq j \leq v \leq k \leq m$, applicando ripetutamente (3.8)

$$\frac{\rho_{j,k}}{\rho_{v,k}} = \frac{\rho_{j,k-1}}{\rho_{v,k-1}} = \dots = \frac{\rho_{j,v}}{\rho_{v,v}} = \rho_{j,v}$$

perché $\rho_{v,v} = 1$. Dunque (3.8) implica che per $1 \leq j \leq v \leq k \leq m$

$$\rho(j, k) = \rho(j, v)\rho(v, k). \quad (3.9)$$

D'altronde (3.8) è un caso particolare di (3.9), perciò il vettore (X_1, \dots, X_m) è di Markov se e solo se per ogni terna di indici $1 \leq j \leq v \leq k \leq m$ vale (3.9). \square

3.3.2 Processi gaussiani di Markov

Supponiamo che $\{X(t)\}_{t \in T}$ sia un processo gaussiano centrato, a variabili in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Indichiamo anche in questo caso con ρ la funzione

$$\begin{aligned} \rho : T \times T &\rightarrow [0, 1] \\ (t_1, t_2) &\mapsto \frac{\text{cov}\{X(t_1), X(t_2)\}}{\sqrt{\text{Var}X(t_1)\text{Var}\{X(t_2)\}}}. \end{aligned}$$

Dalla proposizione 3.3.2 e dalla definizione di processo di Markov segue subito che $\{X(t)\}_{t \in T}$ è di Markov se e solo se, per ogni $t_1 \leq t_2 \leq t_3 \in T$,

$$\rho(t_1, t_3) = \rho(t_1, t_2)\rho(t_2, t_3). \quad (3.10)$$

Supponiamo che il processo sia stazionario, con funzione di covarianza Γ , allora

$$\rho(t_1, t_2) = \frac{\Gamma(|t_1 - t_2|)}{\Gamma(0)},$$

è una funzione del valore assoluto della differenza $|t_1 - t_2|$. Nel caso di processi gaussiani stazionari, vale la seguente caratterizzazione.

Proposizione 3.3.3. *Supponiamo che $\{X(t)\}_{t \in T}$ sia un processo gaussiano stazionario, con funzione di covarianza Γ . Condizione necessaria e sufficiente affinché $X(t)$ sia di Markov è che:*

- se $T = \mathbb{Z}$, per ogni $t \in T$:

$$\Gamma(|t|) = \Gamma(0)\rho(1)^{|t|}.$$

- se $T = \mathbb{R}$, nel caso la funzione di covarianza non sia identicamente nulla, esiste $\lambda \geq 0$ tale che, per ogni $t \in T$

$$\Gamma(|t|) = \Gamma(0)e^{-\lambda t}.$$

Dimostrazione. Nel caso $T = \mathbb{Z}$, se vale

$$\Gamma(t) = \Gamma(0)\rho(1)^{|t|}$$

è verificata la condizione (3.10) e dunque il processo è di Markov. Viceversa verifichiamo per induzione su $|t|$ che (3.10) implichi una simile parametrizzazione della funzione di covarianza. Se $|t| = 0$ o $|t| = 1$ la tesi è verificata

banalmente; supponiamo che sia verificata per $|t| = k$, dimostriamola per $|t| = k + 1$. Per (3.10) abbiamo che

$$\Gamma(k + 1) = \Gamma(0)\rho(k + 1) = \Gamma(0)\rho(1)^{k+1}.$$

Consideriamo ora il caso reale; in questo caso la dimostrazione segue dal lemma 3.3.2, che verrà dimostrato subito sotto. \square

Lemma 3.3.2. *Sia u una funzione*

$$u : \mathbb{R}^+ \rightarrow \mathbb{R}$$

tale che $u(t) \leq 1$ per ogni $t \in \mathbb{R}^+$ e

$$u(t_1 + t_2) = u(t_1)u(t_2) \tag{3.11}$$

per ogni $t_1, t_2 \in \mathbb{R}$. Allora o $u = 0$ oppure esiste $\lambda \in \mathbb{R}^+$ tale che $u = e^{-\lambda t}$.

Dimostrazione. Supponiamo che esista $\bar{t} \in \mathbb{R}^+$ tale che $u(\bar{t}) = 0$. Allora per ogni $z > \bar{t}$

$$u(z) = u(\bar{t})u(z - \bar{t}) = 0.$$

Inoltre per ogni $n \in \mathbb{N}$ vale $u(\bar{t}) = u^n(\bar{t}2^{-n})$, dunque per ogni $n \in \mathbb{N}$

$$u(\bar{t}2^{-n}) = 0.$$

Poiché la successione $\bar{t}2^{-n}$ è infinitesima, dato $t > 0$ esiste $n_t \in \mathbb{N}$ tale che

$$\bar{t}2^{-n_t} < t,$$

dunque per quanto detto sopra $u(t) = 0$ per ogni $t > 0$.

Supponiamo viceversa che $u(t) > 0$ per ogni $t > 0$. Esiste $\lambda \in \mathbb{R}^+$ tale che

$$e^{-\lambda} = u(1).$$

Consideriamo la funzione

$$v(t) = u(t)e^{\lambda t} :$$

essa verifica ancora (3.11), inoltre l'ipotesi di limitatezza di u implica che

$$v(t) \leq e^{\lambda t}$$

e $v(1) = 1$. Consideriamo $a, b \in \mathbb{N}$, da (3.11) applicata a v , segue immediatamente che

$$v\left(\frac{a}{b}\right) = v(1)^{\frac{a}{b}},$$

dunque se $q \in \mathbb{Q}$, $v(q) = 1$. Supponiamo che esista $w \in \mathbb{R}^+$ tale che $v(w) \neq 1$. Se $v(w) > 1$ la funzione v non è limitata, perché

$$v(nw) = v(w)^n.$$

Se invece fosse $v(w) < 1$, allora scelto un razionale $q > w$, avremmo che

$$v(q - w) = \frac{v(q)}{v(w)} > 1,$$

per cui ci si riconduce al caso precedente. Quindi, l'ipotesi che $v(w) \neq 1$ implica che v non sia limitata.

Dal momento che \mathbb{Q} è denso in \mathbb{R} , fissati $\epsilon > 0$ e $t > 0$ esiste $t_\epsilon \in \mathbb{R}^+$ con $t_\epsilon < \epsilon$ e $t - t_\epsilon \in \mathbb{Q}$. Dunque si ha:

$$v(t) = v(t - t_\epsilon)v(t_\epsilon) = 1 \cdot v(t_\epsilon) = v(t_\epsilon).$$

Perciò se v non è limitata, non è limitata in alcun intorno contenente l'origine, il che è assurdo, perché

$$v(t) \leq e^{\lambda t}.$$

Per ogni $t > 0$ deve essere $v(t) = 1$ e quindi $u(t) = e^{\lambda t}$. □

Capitolo 4

Ricerca della trasformazione per la gaussianità

In questo capitolo cominciamo l'analisi statistica del modello proposto da [7] sulla base dei dati a nostra disposizione. Supponiamo che $\{X(t)\}_{t \in \mathbb{R}^+}$ sia il processo che modella il decadimento del numero di linfociti $CD4+$ per mm^3 di sangue nei soggetti sieronegativi. Gli autori di [7] ipotizzano che questo processo sia log-gaussiano e stazionario, ossia che il processo

$$\{\log(X(t))\}_{t \in \mathbb{R}^+}$$

sia un processo gaussiano stazionario. Il primo passo per verificare il modello è sottoporre a test questa ipotesi. Non indagheremo la stazionarietà del processo: assumendo la stazionarietà come ipotesi di partenza, intendiamo verificare se $\log X(t)$ sia gaussiano o meno.

Ogni soggetto del nostro campione è stato richiamato a 45 visite, con un intervallo di sei mesi tra una visita e la successiva. Dall'ipotesi di stazionarietà del processo segue che, se dividiamo le misure a nostra disposizione in base al numero delle visite, otteniamo 45 campioni a variabili indipendenti e identicamente distribuite, con legge uguale a quella di $X(t)$. Sottoponiamo ogni campione a quattro test di gaussianità: i test di Anderson-Darling, di Cramer-Von Mises, di Pearson e di Kolmogorov-Smirnov. Per approfondimenti relativi a questi test, vedi [44], [46]. Nella tabella (4.1) sono riportati i *p-values* del primo campione (i risultati ottenuti per gli altri campioni sono analoghi).

Al livello di significatività 0.1, tutti i test condotti respingono l'ipotesi di gaussianità.

Tabella 4.1: P-values dei test di gaussianità sul primo campione

Anderson-Darling	Cramer-von Mises	Pearson	Kolmogorov-Smirnov
6.08 E-07	1.61E-05	0.07161	0.0007277

Christopher Cox, in un'e-mail, ci ha suggerito che la trasformazione da apportare al processo $X(t)$ può non essere logaritmica e va stimata sperimentalmente a partire dal campione. Fra le possibili trasformazioni, Cox propone la radice quadrata o la radice cubica. Anche con queste trasformazioni, tuttavia, i test condotti sul nostro campione hanno respinto l'ipotesi di gaussianità.

In questo capitolo, presentiamo il metodo utilizzato più frequentemente in letteratura per la ricerca della trasformazione (metodo di Box Cox), e ne proponiamo uno alternativo, basato sui *p-values* del test di Kolmogorov-Smirnov. Nel caso in esame, i due metodi portano a risultati equivalenti. Entrambi i metodi si basano sull'ipotesi che esista un'opportuna trasformazione T tale che il processo $T(X(t))$ sia gaussiano (in realtà ci sarà bisogno di opportune modifiche alla gaussianità): nel paragrafo 4.1 studieremo i rapporti esistenti fra il processo $X(t)$ e il processo $T(X(t))$ per un'opportuna classe parametrizzata di trasformazioni.

4.1 Le distribuzioni Box Cox

Lo scopo del capitolo è determinare una trasformazione T tale che il processo trasformato

$$\{T(X(t))\}_{t \in T}$$

sia gaussiano. In questa sezione introdurremo la famiglia di trasformazioni di Box Cox (il metodo di Box Cox, trattato in 4.2, sarà uno dei due metodi utilizzati per la scelta della trasformazione) e studieremo i rapporti esistenti fra il processo $\{X(t)\}_{t \in T}$ e il processo trasformato $\{f(X(t))\}_{t \in T}$.

Consideriamo, al variare di $\lambda \in \mathbb{R}$, la famiglia di funzioni definite su \mathbb{R}^+

$$f_\lambda(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log x & \text{se } \lambda = 0. \end{cases} \quad (4.1)$$

Esse si dicono trasformazioni di Box Cox, o trasformazioni di potenze. Vogliamo analizzare la distribuzione di una variabile aleatoria reale X , quasi certamente positiva, tale che $f_\lambda(X)$ sia una normale (o quantomeno una normale troncata, che sarà definita in seguito, dato che per $\lambda \neq 0$ la normalità non sarebbe possibile, essendo le immagini delle funzioni f_λ se $\lambda \neq 0$ inferiormente o superiormente limitate).

Separiamo il caso $\lambda = 0$ dagli altri; in tal caso non si parla di distribuzioni di Box Cox ma di log-gaussiane.

4.1.1 Le distribuzioni log-gaussiane

Definizione 4.1.1. *Sia X una variabile aleatoria definita su $(\Omega, \mathcal{F}, \mathbb{P})$, quasi certamente positiva. Se*

$$\log X \sim N(\mu, \sigma^2)$$

si dice che X ha distribuzione log-gaussiana di parametri μ e σ^2 e si scrive $X \sim \log N(\mu, \sigma^2)$.

Supponiamo che X sia una variabile log-gaussiana, possiamo calcolarne densità e funzione di ripartizione.

Proposizione 4.1.1. *Sia X una variabile aleatoria con $X \sim \log N(\mu, \sigma^2)$, se indichiamo con Ψ la funzione di ripartizione di una normale standard, la funzione di ripartizione F di X è, per $t > 0$:*

$$F(t) = \Psi\left(\frac{\log t - \mu}{\sigma}\right).$$

La distribuzione di X ha densità $f(x)$ rispetto alla misura di Lebesgue:

$$f(x) = I_{\mathbb{R}^+}(x) \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\log x - \mu)^2}{\sigma^2}},$$

dove $I_{\mathbb{R}^+}$ è l'indicatrice di $\mathbb{R}^+ = (0, \infty)$.

Dimostrazione. Ricavare la funzione di ripartizione è un semplice calcolo: se $t > 0$

$$\mathbb{P}\{X \leq t\} = \mathbb{P}\{\log X \leq \log t\} = \Psi\left(\frac{\log t - \mu}{\sigma}\right).$$

La funzione

$$x \mapsto \log(x)$$

è un diffeomorfismo di \mathbb{R}^+ su \mathbb{R} , quindi

$$f(x) = g(\log x) \left| \frac{d \log x}{dx} \right| = g(\log x) \frac{1}{x}$$

dove g indica la densità di una variabile normale con media μ e varianza σ^2 . \square

La seguente proposizione mostra che una variabile log-gaussiana ha momenti di ogni ordine.

Proposizione 4.1.2. *Sia $X \sim \log N(\mu, \sigma^2)$, allora X ammette momenti di ogni ordine e*

$$\mathbb{E}\{X^r\} = e^{r\mu + \frac{r^2\sigma^2}{2}}.$$

Dimostrazione. Per la definizione di variabile log-gaussiana, sappiamo che $\log(X) \sim N(\mu, \sigma^2)$, dunque

$$\mathbb{E}\{X^r\} = \mathbb{E}\{e^{rY}\}$$

dove $Y \sim N(\mu, \sigma^2)$. Questo valore atteso corrisponde alla funzione generatrice di momenti della variabile Y , dunque è finito e dato da:

$$e^{r\mu + \frac{r^2\sigma^2}{2}}.$$

\square

La definizione di distribuzione log-gaussiana si estende facilmente al caso vettoriale.

Definizione 4.1.2. *Sia (X_1, \dots, X_m) un vettore aleatorio definito su $(\Omega, \mathcal{F}, \mathbb{P})$; supponiamo che*

$$(\log(X_1), \dots, \log(X_m)) \sim N_m(\mathbf{m}, Q)$$

allora il vettore (X_1, \dots, X_m) si dice log-gaussiano e si indica con

$$(X_1, \dots, X_m) \sim \log N_m(\mathbf{m}, Q).$$

Supponiamo che $(X_1, \dots, X_m) \sim \log N_m(\mathbf{m}, Q)$ e che la matrice Q non sia singolare, allora la distribuzione di (X_1, \dots, X_m) è dotata di una densità $f(x_1, \dots, x_m)$ rispetto alla misura di Lebesgue m -dimensionale, data da

$$\prod_{i=1}^m x_i^{-1} \frac{1}{(2\pi)^{\frac{m}{2}} \det Q} e^{-(\log \mathbf{X} - \mathbf{m})^t Q^{-1} (\log \mathbf{X} - \mathbf{m})},$$

dove $\log \mathbf{X} = (\log X_1, \dots, \log X_n)$.

Proposizione 4.1.3. *Sia $(X_1, X_2) \sim \log N_2(\mathbf{m}, Q)$, allora*

$$\text{cov}\{X_1, X_2\} = e^{\mathbf{m}_1 + \mathbf{m}_2 + \frac{1}{2}(q_{11} + q_{2,2})} (e^{q_{1,2}} - 1).$$

Dimostrazione. Per definizione di vettore log-gaussiano

$$(\log X_1, \log X_2) \sim N_2(\mathbf{m}, Q)$$

dunque

$$\mathbb{E}\{X_1 X_2\} = \mathbb{E}\{e^{Y_1} e^{Y_2}\}$$

dove $(Y_1, Y_2) \sim N_2(\mathbf{m}, Q)$. Il valore atteso è la funzione generatrice di momenti di un vettore gaussiano:

$$\mathbb{E}\{X_1 X_2\} = e^{\mathbf{m}_1 + \mathbf{m}_2 + \frac{1}{2}(q_{11} + q_{2,2}) + q_{1,2}}.$$

Dunque

$$\text{cov}\{X_1, X_2\} = \mathbb{E}\{X_1 X_2\} - \mathbb{E}\{X_1\} \mathbb{E}\{X_2\} = e^{\mathbf{m}_1 + \mathbf{m}_2 + \frac{1}{2}(q_{11} + q_{2,2})} (e^{q_{1,2}} - 1).$$

□

Indichiamo, infine, che cosa è un processo stocastico log-gaussiano.

Definizione 4.1.3. *Sia $\{X(t)\}_{t \in T}$ un processo stocastico. Esso si dice log-gaussiano se il processo*

$$\{\log X(t)\}_{t \in T}$$

è un processo gaussiano.

Supponiamo che $\{X(t)\}_{t \in T}$ sia un processo log-gaussiano stazionario e che $\log X(t)$ abbia media μ e funzione di covarianza Γ , per le proposizioni 4.1.2 e 4.1.3 media e funzione di covarianza di $X(t)$ sono

$$\mathbb{E}\{X(t)\} = e^{\mu + \frac{\Gamma(0)}{2}} \quad \text{cov}\{X(t)X(t+h)\} = e^{\mu + \frac{\Gamma(0)}{2}} (e^{\Gamma(h)} - 1).$$

4.1.2 Le distribuzioni di Box Cox con $\lambda \neq 0$

Diamo la definizione di distribuzione gaussiana, o normale, troncata.

Definizione 4.1.4. *Sia X una variabile aleatoria reale, con $X \sim N(\mu, \sigma^2)$. Sia $E = [a, \infty)$ o $E = (-\infty, b]$ una semiretta. Si dice distribuzione gaussiana troncata ad E la legge della variabile aleatoria X condizionale all'evento $\{X \in E\}$. Essa si indica con $TN(\mu, \sigma^2, E)$.*

Supponiamo che $E = [a, \infty)$, con $a \in \mathbb{R}$. Indicheremo con z la funzione che standardizza la variabile $N(\mu, \sigma^2)$:

$$z(t) = \frac{t - \mu}{\sigma}.$$

è facile ricavare la funzione di ripartizione F di una variabile $X \sim TN(\mu, \sigma^2, E)$. Se Ψ è la funzione di ripartizione di una variabile normale standard, si ha infatti

$$F(t) = \begin{cases} \frac{\Psi(z(t)) - \Psi(z(a))}{1 - \Psi(z(a))} & \text{se } t \geq a \\ 0 & \text{se } t < a. \end{cases}$$

La densità f di X rispetto alla misura di Lebesgue è:

$$f(x) = I_E(x) \frac{1}{1 - \Psi(z(a))} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-z(x)^2\}.$$

dove I_E è l'indicatrice di E . Su questo insieme la densità coincide, a meno di una costante moltiplicativa, con la densità di una variabile normale: segue che le variabili normali troncate ammettono momento di ogni ordine.

Il caso $E = (-\infty, b]$ è del tutto analogo: procedendo come in precedenza si ricava che la funzione di ripartizione F e la densità f rispetto alla misura di Lebesgue di una variabile con legge $TN(\mu, \sigma^2, E)$ sono rispettivamente:

$$F(t) = \begin{cases} \frac{\Psi(z(t))}{\Psi(z(b))} & \text{se } t \leq b \\ 1 & \text{se } t > b \end{cases} \quad f(x) = I_E(x) \frac{1}{\Psi(z(b))} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-z(x)^2\}.$$

Possiamo definire le distribuzioni di Box Cox.

Definizione 4.1.5. *Sia $\lambda > 0$ e X una variabile aleatoria reale, con*

$$X \sim TN(\mu, \sigma^2, [-\frac{1}{\lambda}, \infty));$$

la distribuzione della variabile aleatoria $(\lambda X + 1)^{\frac{1}{\lambda}}$ si dice distribuzione di Box Cox di parametri μ , σ^2 e λ , ed è indicata con $\text{BC}(\mu, \sigma^2, \lambda)$.

Viceversa, se $\lambda < 0$ e

$$X \sim \text{TN}(\mu, \sigma^2, (-\infty, \frac{1}{\lambda}]);$$

la distribuzione della variabile aleatoria $(\lambda X + 1)^{\frac{1}{\lambda}}$ si dice di Box Cox di parametri μ , σ^2 e λ e si indica con $\text{BC}(\mu, \sigma^2, \lambda)$.

Supponiamo $\lambda > 0$ (il caso $\lambda < 0$ è analogo): la funzione

$$x \mapsto (\lambda x + 1)^{\frac{1}{\lambda}},$$

definita sul dominio $[-\frac{1}{\lambda}, \infty)$, è a valori positivi, dunque le distribuzioni di Box Cox sono concentrate su \mathbb{R}^+ . Inoltre essa è l'inversa della trasformazione di Box Cox f_λ , che è stata definita in (4.1).

Proposizione 4.1.4. *Sia $X \sim \text{BC}(\mu, \sigma^2, \lambda)$, poniamo*

$$c = \frac{1}{\lambda\sigma} + \frac{m}{\sigma}$$

e indichiamo con Ψ la funzione di ripartizione di una variabile normale standard. Definita la funzione

$$z(x) = \frac{x - \mu}{\sigma},$$

la funzione di ripartizione F di X è:

$$F(t) = \begin{cases} \frac{\Psi(z(t)) - \Psi(-c)}{\Psi(c)} & \text{se } \lambda > 0 \\ \frac{\Psi(z(t))}{\Psi(-c)} & \text{se } \lambda < 0. \end{cases}$$

Dimostrazione. La dimostrazione è un semplice calcolo; considerando solo il caso di $\lambda > 0$:

$$\mathbb{P}\{X \leq t\} = \mathbb{P}\{f_\lambda(X) \leq \frac{t^\lambda - 1}{\lambda}\}.$$

Dunque, siccome per definizione di distribuzione di Box Cox

$$T_\lambda(X) \sim \text{TN}(\mu, \sigma^2, [-\frac{1}{\lambda}, \infty)),$$

ricordando quale sia la funzione di ripartizione di una distribuzione normale troncata, si ha:

$$F(t) = \frac{\Psi(z(t)) - \Psi(-c)}{\Psi(c)}.$$

□

Le trasformazioni f_λ definite in (4.1) sono diffeomorfismi a valori in \mathbb{R}^+ . Possiamo quindi ricavare anche le densità di variabili con distribuzioni di Box Cox rispetto alla misura di Lebesgue:

$$\frac{d\text{BC}(\mu, \sigma^2, \lambda)}{d\lambda}(x) = I_{\mathbb{R}^+}(x) \frac{1}{K(\lambda)\sqrt{2\pi\sigma^2}} x^{\lambda-1} \exp\left\{-\frac{(f_\lambda(x) - \mu)^2}{\sigma^2}\right\},$$

dove $K(\lambda)$ è la funzione

$$K(\lambda) = \begin{cases} \Psi(c) & \text{se } \lambda > 0 \\ \Psi(-c) & \text{se } \lambda < 0 \end{cases}$$

e c è già stato definito nella proposizione 4.1.4. Nella proposizione seguente calcoliamo i momenti di una variabile di Box Cox in condizioni particolari.

Proposizione 4.1.5. *Sia X una variabile aleatoria con*

$$X \sim \text{BC}(\mu, \sigma^2, \lambda).$$

Supponiamo che $\lambda > 0$ e che $\mu > -\frac{1}{\lambda}$, allora X ammette momento di ogni ordine. Indicata con Z la variabile aleatoria

$$\frac{f_\lambda(X) - \mu}{\sigma},$$

il momento r -esimo di X è

$$\mathbb{E}\{X^r\} = \sum_{i=0}^{\infty} \frac{1}{i!} g^i(\mu) \sigma^i \mathbb{E}\{Z^i\},$$

dove $g(t) = (\lambda t + 1)^{\frac{r}{\lambda}}$.

Dimostrazione. Dimostriamo innanzitutto che la variabile aleatoria X ammette momenti di ogni ordine. Dato $r \in \mathbb{N}$

$$\mathbb{E}\{|X|^r\} = \mathbb{E}\{|T_\lambda^{-1}(f_\lambda(X))|^r\} = \mathbb{E}\{(\lambda T_\lambda(X) + 1)^{\frac{r}{\lambda}}\} \quad (4.2)$$

Per la definizione di distribuzione di Box Cox, $T_\lambda(X)$ è una variabile gaussiana troncata, perciò anche $(\lambda T_\lambda(X) + 1)$ è una variabile gaussiana troncata. Allora

$$\mathbb{E}\{(\lambda T_\lambda(X) + 1)^{\frac{r}{\lambda}}\} < \infty.$$

Per il calcolo del momento, introduciamo la funzione

$$g : \left(-\frac{1}{\lambda}, +\infty\right) \rightarrow \mathbb{R}^+ \quad (4.3)$$

$$t \mapsto (\lambda t + 1)^{\frac{r}{\lambda}}.$$

Essa può essere sviluppata in serie di Taylor, con punto iniziale in $\mu > -\frac{1}{\lambda}$:

$$g(y) = \sum_{i=0}^{\infty} \frac{g^{(i)}(\mu)}{i!} (y - \mu)^i, \quad (4.4)$$

e la serie converge uniformemente su tutto il dominio a $g(y)$. Possiamo quindi integrare per serie, ottenendo

$$\mathbb{E}\{X^r\} = \sum_{i=0}^{\infty} \frac{1}{i!} g^{(i)}(\mu) \sigma^i \mathbb{E}\{Z^i\}$$

con $Z = \frac{f_\lambda(X) - \mu}{\sigma}$. □

In particolare, se approssimiamo la distribuzione di Y ad una distribuzione normale $N(m, \sigma^2)$, si ha

$$\mathbb{E}\{(Y - \mu)^i\} = \begin{cases} \frac{\sigma^i i!}{2^{\frac{i}{2}} \frac{i!}{2!}} & \text{per } i \text{ dispari} \\ 0 & \text{per } i \text{ pari} \end{cases}$$

e dunque

$$\mathbb{E}\{X^r\} = \sum_{i=0}^{\infty} \frac{\sigma^i}{2^{\frac{i}{2}} \frac{i!}{2!}} g^{(i)}(\mu).$$

Procedendo in maniera analoga a quanto fatto nel caso unidimensionale, si possono definire le distribuzioni di Box Cox multidimensionali. Definiamo dapprima le distribuzioni gaussiane multidimensionali troncate.

Definizione 4.1.6. Sia X un vettore gaussiano a valori in \mathbb{R}^n , con vettore delle medie \mathbf{m} e matrice di covarianza Q ; sia $E \subset \mathbb{R}^n$ un plurirettangolo con $\mathbb{P}\{X \in E\} \neq 0$, la legge di X condizionata a $X \in E$ si dice gaussiana multidimensionale troncata e si indica con $\text{TN}_n(\mathbf{m}, Q, E)$.

Analogamente al caso dei vettori gaussiani, se

$$X = (X_1, \dots, X_n) \sim \text{TN}_n(\mathbf{m}, Q, E),$$

ogni vettore

$$(X_{k_1}, \dots, X_{k_m}) \quad \text{con } k_1, \dots, k_m \in \{1, \dots, n\}$$

ha distribuzione normale troncata.

Supponiamo che X sia un vettore gaussiano con media \mathbf{m} e matrice di covarianza Q non singolare, allora la misura di probabilità $\text{TN}_n(\mathbf{m}, Q, E)$ ammette densità rispetto alla misura di Lebesgue n -dimensionale, data da:

$$f(\mathbf{x}) = I_E(\mathbf{x}) \frac{1}{\mathbb{P}\{X \in E\}} \frac{1}{\sqrt{2\pi \det Q}} \exp\left\{-\frac{1}{2} \langle \mathbf{x} - \mathbf{m}, Q^{-1}(\mathbf{x} - \mathbf{m}) \rangle\right\}.$$

Possiamo definire le distribuzioni di Box Cox multidimensionali.

Definizione 4.1.7. Supponiamo che Y sia un vettore aleatorio a valori in \mathbb{R}^n e $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R} \setminus \{0\}$ tali che

$$Y \sim \text{TN}(\mathbf{m}, Q, C(\lambda_1) \times \dots \times C(\lambda_n)),$$

dove $C(\lambda_i)$ è definito da:

$$C(\lambda_i) = \begin{cases} [-\frac{1}{\lambda_i}, +\infty) & \text{se } \lambda_i > 0 \\ (-\infty, -\frac{1}{\lambda_i}] & \text{se } \lambda_i < 0. \end{cases}$$

La distribuzione del vettore aleatorio (X_1, \dots, X_n)

$$X_i = (\lambda_i Y_i + 1)^{\frac{1}{\lambda_i}},$$

si dice Box Cox multidimensionale, con parametri \mathbf{m} , Q e $\lambda = (\lambda_1, \dots, \lambda_n)$, e si indica con

$$\text{BC}_n(\mathbf{m}, Q, \lambda).$$

Per quanto detto circa i vettori con legge gaussiana troncata, se $X = (X_1, \dots, X_m)$ ha distribuzione di Box Cox multidimensionale, qualunque vettore da esso estratto avrà legge di Box Cox multidimensionale. Supponiamo che la matrice Q sia non singolare, allora la misura di probabilità $BC_n(\mathbf{m}, \Gamma, \lambda)$ ammette densità rispetto alla misura di Lebesgue n -dimensionale:

$$\frac{dBC_n(\mathbf{m}, Q, \lambda)}{d\lambda^n} = \frac{1}{K} \prod_{i=1}^n z_i^{\lambda_i-1} f(T_{\lambda_1}(x_1), \dots, T_{\lambda_n}(x_n)),$$

dove K è un'opportuna costante di normalizzazione e $f(y_1, \dots, y_n)$ indica la densità di una distribuzione gaussiana multidimensionale di media \mathbf{m} e matrice di covarianza Q .

Proposizione 4.1.6. *Sia $X = (X_1, X_2)$ un vettore aleatorio, con*

$$X \sim BC_2(\mathbf{m}, Q, (\lambda_1, \lambda_2)).$$

Supponiamo che $\lambda_1, \lambda_2 > 0$ e che $\mathbf{m} \in [-\frac{1}{\lambda_1}, \infty) \times [-\frac{1}{\lambda_2}, \infty)$, allora

$$\mathbb{E}\{X_1 X_2\} = \sum_{i,j=0}^{\infty} \frac{1}{i!j!} g_1^i(m_1) g_2^j(m_2) \rho_{1,2},$$

dove le funzioni g_i sono definite da

$$g_i = T_{\lambda_i}^{-1}$$

e $\rho_{1,2}$ è il momento incrociato $\mathbb{E}\{(Y_1 - m_1)^i (Y_2 - m_2)^j\}$ di un vettore

$$(Y_1, Y_2) \sim TN_2(\mathbf{m}, Q, [-\frac{1}{\lambda_1}, \infty) \times [-\frac{1}{\lambda_2}, \infty)).$$

Dimostrazione. Per la definizione di distribuzione multidimensionale di Box Cox

$$(T_{\lambda_1}(X_1), T_{\lambda_2}(X_2)) \sim TN_2(\mathbf{m}, Q, [-\frac{1}{\lambda_1}, \infty) \times [-\frac{1}{\lambda_2}, \infty)),$$

dunque

$$\mathbb{E}\{X_1 X_2\} = \mathbb{E}\{g_1(Y_1) g_2(Y_2)\}$$

dove (Y_1, Y_2) è un vettore aleatorio con distribuzione

$$(Y_1, Y_2) \sim \text{TN}(\mathbf{m}, Q, [-\frac{1}{\lambda_1}, \infty) \times [-\frac{1}{\lambda_2}, \infty)).$$

Consideriamo la funzione

$$\begin{aligned} [-\frac{1}{\lambda_1}, \infty) \times [-\frac{1}{\lambda_2}, \infty) &\rightarrow \mathbb{R}^+ \\ (x, y) &\mapsto g_1(x_1)g_2(x_2) \end{aligned}$$

siccome g_1 e g_2 sono sviluppabili in serie di potenze attorno a qualsiasi punto del loro dominio, lo è anche il prodotto g_1g_2 . Scegliamo come punto iniziale (m_1, m_2) :

$$g_1(x)g_2(y) = \sum_{i,j=0}^{\infty} \frac{1}{i!j!} g_1^i(m_1)g_2^j(m_2)(x - m_1)^i(y - m_2)^j$$

e la convergenza è uniforme. Quindi

$$\mathbb{E}\{X_1X_2\} = \mathbb{E}\{g_1(Y_1)g_2(Y_2)\} = \sum_{i,j=0}^{\infty} \frac{1}{i!j!} g_1^i(m_1)g_2^j(m_2)\mathbb{E}\{(Y_1 - m_1)^i(Y_2 - m_2)^j\}.$$

□

Le definizioni e le dimostrazioni date possono essere estese facilmente al caso di processi stocastici: dato un processo $\{X(t)\}_{t \in T}$, con $T \subset \mathbb{R}$, esso si dirà di Box Cox se comunque scelti $t_1, \dots, t_n \in T$, il vettore aleatorio $(X(t_1), \dots, X(t_n))$ è di Box Cox. Grazie a quanto dimostrato precedentemente sappiamo ricavare le relazioni fra le funzioni di media e covarianza di $\{X(t)\}_{t \in T}$ e quelle del processo trasformato $\{f_\lambda(X(t))\}_{t \in T}$.

4.2 Il metodo di Box Cox

4.2.1 Costruzione del metodo

In questo paragrafo riportiamo la costruzione del metodo di Box Cox esposta dagli stessi D. R. Cox e G. E. P. Box in [19]. Una critica e delle possibili giustificazioni a questo metodo saranno trattate nel paragrafo successivo.

Supponiamo di avere un campione (X_1, \dots, X_n) a variabili indipendenti ed identicamente distribuite definito sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Supponiamo inoltre che le variabili aleatorie X_i siano, almeno quasi certamente, positive. Consideriamo, al variare di $\lambda \in \mathbb{R}$, la famiglia di trasformazioni di Box Cox:

$$T_\lambda(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log y & \text{se } \lambda = 0 \end{cases}$$

Se esiste $\lambda \in \mathbb{R}$ tale che $f_\lambda(X_i) \sim N(m, \sigma)^1$, la densità del vettore aleatorio $(T_\lambda X_1), \dots, T_\lambda(X_n)$ rispetto alla misura di Lebesgue n -dimensionale è:

$$L(T_\lambda(x_1), \dots, T_\lambda(x_n)) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{\sum_{i=1}^n \frac{(T_\lambda(x_i) - m)^2}{\sigma^2}\right\}.$$

Una volta definito opportunamente il codominio, la funzione

$$(x_1, \dots, x_n) \mapsto (T_\lambda(x_1), \dots, T_\lambda(x_n))$$

è un diffeomorfismo su $\mathbb{R}^+ \times \dots \times \mathbb{R}^+$; sia $J_\lambda(x_1, \dots, x_n)$ la sua matrice jacobiana. La funzione di verosimiglianza del modello statistico originale sarà data da:

$$L(\lambda, m, \sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{\sum_{i=1}^n (T_\lambda(x_i) - m)^2\right\} |\det J_\lambda(x_1, \dots, x_n)|.$$

Vogliamo ottenere uno stimatore di massima verosimiglianza per i parametri μ , λ e σ^2 .

Per farlo notiamo che, fissato λ , massimizzare $L(\lambda, m, \sigma^2)$ è equivalente a massimizzare la verosimiglianza del campione

$$(T_\lambda(X_1), \dots, T_\lambda(X_n))$$

a variabili indipendenti ed identicamente distribuite, con legge $N(m, \sigma^2)$. Sono noti gli stimatori

$$\hat{m}_\lambda(f_\lambda(X_1), \dots, T_\lambda(X_n)) = \frac{1}{n} \sum_{i=1}^n T_\lambda(X_i)$$

¹Riportiamo il metodo di Box Cox solo per determinare la normalità del campione, anche se nell'articolo [19] si affronta il problema più completo di garantire le ipotesi di gaussianità e di uguaglianza delle varianze dei residui nelle equazioni di regressione.

$$\hat{\sigma}_\lambda^2(T_\lambda(X_1), \dots, T_\lambda(X_n)) = \frac{1}{n} \sum_{i=1}^n (T_\lambda(X_i) - \hat{m}_\lambda)^2.$$

Li abbiamo indicati con \hat{m}_λ e $\hat{\sigma}_\lambda^2$ per ricordare che essi non sono gli stimatori di verosimiglianza globali, ma quelli che si ottengono fissando λ . Inoltre

$$|\det J_\lambda(x_1, \dots, x_n)| = \prod_{i=1}^n x_i^{\lambda-1}.$$

Sostituendo gli stimatori nell'equazione di verosimiglianza, otteniamo il cosiddetto profilo della verosimiglianza

$$L_P(\lambda) = L(\lambda, \hat{m}_\lambda, \hat{\sigma}_\lambda^2)$$

e

$$l_P = \log L_P(\lambda) = K - \frac{n}{2} \log(\hat{\sigma}_\lambda^2) + (\lambda - 1) \sum_{i=1}^n \log(x_i).$$

Se il massimo della funzione di verosimiglianza esiste, esso deve essere nella forma $(\alpha, \hat{m}_\alpha, \hat{\sigma}_\alpha^2)$, dove α è il massimo della funzione $L_P(\lambda)$.

A questo punto lo stimatore di massima verosimiglianza $\hat{\lambda}$ viene, di solito, calcolato numericamente.

Oltre ad uno stimatore puntuale del parametro, possiamo costruire il test di rapporto di verosimiglianza per λ , di ipotesi nulla e alternativa rispettivamente

$$H_0 : \lambda = \lambda_0 \quad H_1 : \lambda \neq \lambda_0.$$

Esso si baserà sulla statistica

$$2(l_P(\hat{\lambda}) - l_P(\lambda_0)).$$

Ricordiamo il teorema di Wilks (per una dimostrazione, vedi [48]):

Teorema 4.2.1. *Consideriamo un modello statistico parametrico*

$$(\Omega, \mathcal{F}, \{\mathbb{P}^\theta\}_{\theta \in \Theta}),$$

con $\Theta \subset \mathbb{R}^n$ aperto e sia $\Theta_0 \subset \Theta$ aperto, tale che Θ_0 sia contenuto in un sottospazio vettoriale di \mathbb{R}^n , di dimensione $m < n$.

Supponiamo che $\{X_n\}_{n \in \mathbb{N}}$ sia una successione di variabili aleatorie reali indipendenti ed identicamente distribuite definite su (Ω, \mathcal{F}) . Sia $f(x_1, \dots, x_n, \theta)$

la verosimiglianza rispetto alla misura di Lebesgue del modello indidotto dal campione X_1, \dots, X_n .

Se per ogni $n \in \mathbb{N}$ è definito uno stimatore di massima verosimiglianza $\hat{\theta}_n$ per θ e la successione $\hat{\theta}_n$ è asintoticamente normale, allora il rapporto di verosimiglianza

$$R_n = 2 \log \frac{\sup_{\theta \in \Theta_0} f(X_1, \dots, X_n, \theta)}{\sup_{\theta \in \Theta} f(X_1, \dots, X_n, \theta)}$$

è tale che:

$$\lim_{n \rightarrow \infty} \mathbb{P}^\theta(R_n > h) = \int_h^\infty f_{\chi_p^2}(x) dx,$$

dove $f_{\chi_p^2}$ indica la densità di una variabile con distribuzione χ_p^2 e $p = n - m$

Secondo il teorema di Wilks, dunque

$$2(l_P(\hat{\lambda}) - l_P(\lambda_0)) \xrightarrow{\mathcal{L}} Y,$$

dove $Y \sim \chi^2(1)$. Abbiamo a disposizione dunque un test di livello asintotico α , con $\alpha \in (0, 1)$ per l'ipotesi nulla $\lambda = \lambda_0$ contro l'alternativa $\lambda \neq \lambda_0$: come è stato spiegato in 2.2.3 è possibile costruire degli intervalli di confidenza con livello di significatività $1 - \alpha$ per λ .

4.2.2 Un approccio formale e giustificazioni del metodo

La costruzione del metodo di Box Cox non è formalmente corretta: si possono muovere per lo meno due critiche.

La prima riguarda l'effettiva distribuzione del campione (X_1, \dots, X_n) : affinché siano ben definite le trasformazioni f_λ bisogna assumere che le variabili aleatorie X_i siano quasi certamente positive, ma dunque le variabili trasformate $f_\lambda(X_i)$ sono quasi certamente inferiormente o superiormente limitate, per $\lambda \neq 0$. Questo implica che nessuna $f_\lambda(X_i)$, con $\lambda \neq 0$, possa avere distribuzione gaussiana.

Una seconda critica riguarda l'utilizzo degli stimatori di massima verosimiglianza, che in [19] non sembra supportato da alcuna considerazione circa il modello statistico considerato.

In questo paragrafo ci concentreremo sulla prima critica: mostreremo quali

complicazioni comportsi procedere rigorosamente e accenneremo agli argomenti che Cox stesso ha addotto in favore del suo metodo.

Rigorosamente la distribuzione della variabile trasformata che va considerata non è gaussiana, ma gaussiana troncata. Supponiamo che esista $\lambda \in \mathbb{R}$ tale che $f_\lambda(X)$ abbia distribuzione $\text{TN}(\mu, \sigma^2, C(\lambda))$, dove l'insieme $C(\lambda)$ è definito in questo modo:

$$C(\lambda) = \begin{cases} [-\frac{1}{\lambda}, \infty) & \text{se } \lambda > 0 \\ \mathbb{R} & \text{se } \lambda = 0 \\ (-\infty, -\frac{1}{\lambda}] & \text{se } \lambda < 0 \end{cases}$$

Il nostro scopo è quello di determinare una trasformazione f_λ per cui $f_\lambda(X)$ sia una variabile approssimativamente normale. Ma cosa significa che una variabile ha legge approssimativamente normale?

Fissiamo, a nostro piacere, un $\epsilon > 0$ e, detta F' la funzione di ripartizione di una variabile con distribuzione $\text{TN}(\mu, \sigma^2, C(\lambda))$ e F quella della relativa distribuzione normale $N(\mu, \sigma^2)$, imponiamo che:

$$\sup_{t \in \mathbb{R}} |F'(t) - F(t)| \leq \epsilon. \quad (4.5)$$

Se supponiamo di restringere l'intervallo a $\lambda > 0$, otteniamo che la condizione (4.5) è equivalente a:

$$2 \frac{\Psi((-\frac{1}{\lambda} - \mu)/\sigma)}{1 - \Psi((-\frac{1}{\lambda} - \mu)/\sigma)} \leq \epsilon,$$

dove Ψ è la funzione di ripartizione di una normale standard. Fissati μ e σ^2 , si ha

$$\lim_{\lambda \rightarrow 0} \frac{\Psi((-\frac{1}{\lambda} - \mu)/\sigma^2)}{1 - \Psi((-\frac{1}{\lambda} - \mu)/\sigma^2)} = 0$$

quindi l'insieme dei parametri $A \subset \mathbb{R}^3$ su cui (4.5) è soddisfatta non è vuoto. Proseguendo in maniera analoga a quanto fatto nel paragrafo precedente, tentiamo di derivare degli stimatori di massima verosimiglianza per i parametri, tenuto conto che lo spazio dei parametri è stato modificato in A . Per semplicità supponiamo di avere a disposizione un'unica osservazione: per quanto detto nella sezione 4.1 la densità $f(x)$ di X rispetto alla misura di Lebesgue è

$$f(x) = I_{\mathbb{R}^+}(x) \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{K(\lambda, \mu, \sigma)} e^{-\frac{(f_\lambda(x) - \mu)^2}{\sigma^2}}.$$

In questo caso non è possibile ricavare il profilo dell'equazione di massima verosimiglianza rispetto a λ , perché, come osservato in [11], le equazioni di massima verosimiglianza di variabili gaussiane troncate non portano ad una soluzione in forma chiusa. Bisognerebbe allora procedere ad una stima numerica della massima verosimiglianza: in [21] viene spiegato perché questo problema non è facilmente risolvibile, anche solo nei parametri μ e σ^2 .

In difesa della costruzione originale vi sono molti argomenti, riportati in [15]: Cox sostiene che, anche se non esiste un valore del parametro per cui la variabile trasformata è esattamente una gaussiana, il valore di λ stimato possiede varie proprietà legate ai momenti della variabile trasformata $f_\lambda(X)$, fra cui un coefficiente di asimmetria molto basso.

4.2.3 La scelta della trasformazione

Abbiamo a disposizione 45 campioni a variabili indipendenti e identicamente distribuite, tutti con uguale distribuzione.

Per la determinazione della trasformazione che rende la distribuzione dei dati gaussiana, almeno approssimativamente, applichiamo il metodo di Box Cox ad ogni campione. Nella tabella 4.2 sono riportati le stime di λ relative alle prime cinque visite e il relativo intervallo di confidenza di livello 0.05.

Tabella 4.2: Stime e intervalli di confidenza per λ relativi alle prime visite.

	Visita 1	Visita 2	Visita 3	Visita 4	Visita 5
$\hat{\lambda}$	0.12	0.25	0.27	0.25	0.23
C_I	(0.06,0.18)	(0.18, 0.31)	(0.20, 0.34)	(0.17, 0.33)	(0.14, 0.33)

Si pone il problema di scegliere un unico stimatore relativo a tutte le visite. Proseguiamo in due modi: da principio prendiamo semplicemente la media aritmetica degli stimatori; abbiamo

$$\hat{\lambda}_1 = 0.19.$$

Questo valore è contenuto in 34 dei 45 intervalli di confidenza calcolati. Potremmo anche stimare λ di modo che la stima sia contenuta nel maggior

numero di intervalli di confidenza. In questo caso abbiamo la nuova stima:

$$\hat{\lambda}_2 = 0.15,$$

valore che è interno a 37 dei 45 intervalli di confidenza calcolati.

Si tratta di verificare che questi stimatori siano appropriati e di scegliere il migliore. Per questo ripetiamo i test di normalità sulle variabili aleatorie trasformate nei due modi. Consideriamo solo la prima visita, i p -values ottenuti (riportati nella tabella 4.3) in quest'occasione sono simili a quelli relativi alle altre.

Tabella 4.3: P-values relativi alle due trasformazioni individuate.

	Anderson-Darling	Cramer-von Mises	Pearson	Kolmogorov-Smirnov
$\hat{\lambda}_1$	0.1032	0.1103	0.3865	0.06246
$\hat{\lambda}_2$	0.2201	0.2396	0.3649	0.1404

Accettiamo quindi la stima $\hat{\lambda} = 0.19$. Siccome poi è uso comune (vedi [34]), semplificare la trasformazione ottenuta, porremo $\hat{\lambda} = 0.2$.

4.3 Metodo basato sul test di Kolmogorov-Smirnov

Christopher Cox suggerisce² che la trasformazione di cui abbiamo bisogno vada cercata fra le funzioni

$$t \mapsto t^\alpha \quad \text{con } 0 < \alpha < 1.$$

Lo scopo di questa sezione è proporre un metodo alternativo a quello di Box Cox, con basi matematiche più rigorose. In 4.3.3 i due metodi saranno confrontati in base ai risultati ottenuti sui nostri campioni. Il metodo è basato sul test di Kolmogorov-Smirnov, che descriviamo brevemente nel prossimo paragrafo.

²E-mail a chi scrive

4.3.1 Il test di Kolmogorov-Smirnov.

Siano F e G le funzioni di ripartizione di due distribuzioni di probabilità su $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$; definiamo la loro distanza secondo Kolmogorov-Smirnov come

$$d(F, G) = \sup_{t \in \mathbb{R}} |F(t) - G(t)|,$$

essa è una metrica sull'insieme delle funzioni di ripartizione.

Ricordiamo che dato un campione (X_1, \dots, X_n) di variabili aleatorie definite su $(\Omega, \mathcal{F}, \mathbb{P})$, reali, indipendenti ed identicamente distribuite, si dice *funzione di ripartizione empirica* la famiglia di variabili aleatorie, per $t \in \mathbb{R}$:

$$\hat{F}_n(\omega, t) = \frac{1}{n} \sum_{i=1}^n I\{X_i(\omega) \leq t\}.$$

Fissato $\omega \in \Omega$, la funzione

$$t \mapsto \hat{F}_n(\omega, t)$$

è una funzione di ripartizione di una misura di probabilità discreta su \mathbb{R} . In seguito riportiamo il teorema di Glivenko Cantelli (per una dimostrazione si veda [17]).

Teorema 4.3.1. *Data una successione $\{X_n\}_{n \in \mathbb{N}}$ di variabili aleatorie reali, indipendenti ed identicamente distribuite definite su $(\Omega, \mathcal{F}, \mathbb{P})$ e con funzione di ripartizione F ; sia \hat{F}_n la funzione di ripartizione empirica delle variabili X_1, \dots, X_n , \mathbb{P} quasi certamente:*

$$d(\hat{F}_n, F) \rightarrow 0.$$

La costruzione del test di Kolmogorov Smirnov si basa sul teorema di Glivenko-Cantelli.

Supponiamo di avere una successione di variabili aleatorie $\{X_n\}_{n \in \mathbb{N}}$ reali, indipendenti ed identicamente distribuite; sia F la loro funzione di ripartizione. Data una funzione di ripartizione F_0 , consideriamo il problema di testare l'ipotesi nulla $H_0 : F = F_0$ contro l'alternativa $H_1 : F \neq F_0$.

Per il teorema di Glivenko Cantelli

$$d(\hat{F}_n, F_0) \rightarrow d(F, F_0),$$

dove $d(F, F_0) = 0$ se e solo se $F = F_0$. Consideriamo la variabile aleatoria:

$$T_n = \sqrt{n} \cdot d(\hat{F}_n, F_0).$$

In [37] e [1] si dimostra che la distribuzione di T_n , sotto l'ipotesi nulla $F = F_0$, non dipende da F_0 , purché essa sia continua, e che T_n converge in legge ad una variabile aleatoria con distribuzione continua e strettamente crescente su $(0, +\infty)$; indichiamo con F^* la funzione di ripartizione di tale distribuzione. Il test con regione di rifiuto

$$S_\alpha = \{\omega | T_n(\omega) > F^{*-1}(1 - \alpha)\},$$

ha perciò livello di significatività asintotico α .

Per la costruzione che sarà esposta nel paragrafo 4.3.2 abbiamo bisogno di alcuni risultati sul *p-value* del test di Kolmogorov-Smirnov.

Nel capitolo 2 abbiamo ricordato cos'è il *p-value*: nel caso del test di Kolmogorov-Smirnov esso è dato da

$$p_n = \inf_\alpha \{T_n > F^{*-1}(1 - \alpha)\} = 1 - F^*(T_n).$$

Nel prossimo paragrafo saranno sfruttati alcuni risultati asintotici sulla distribuzione del *p-value* del test di Kolmogorov Smirnov.

Se ipotizziamo che $F = F_0$ sappiamo che

$$T_n \xrightarrow{\mathcal{L}} Y,$$

dove Y è una variabile aleatoria con distribuzione F^* : segue che

$$F^*(T_n) \xrightarrow{\mathcal{L}} U$$

dove U è una variabile aleatoria con distribuzione uniforme su $[0, 1]$. Dunque si ha anche

$$p_n \xrightarrow{\mathcal{L}} U.$$

Se invece $F \neq F_0$, quasi certamente

$$T_n \rightarrow \infty;$$

dunque, quasi certamente, la successione di variabili p_n converge puntualmente alla costante 0.

4.3.2 Un metodo basato sul test di Kolmogorov Smirnov

Sia $\{X_n\}_{n \in \mathbb{N}}$ una successione di variabili indipendenti ed identicamente distribuite, definite sullo spazio $(\Omega, \mathcal{F}, \mathbb{P})$. Supponiamo che X_i abbia legge continua, con funzione di ripartizione strettamente crescente e che essa sia, almeno quasi certamente, positiva. Sia, inoltre, \mathbb{P}_0 una distribuzione di probabilità concentrata su \mathbb{R}^+ , con funzione di ripartizione F_0 .

Supporremo che esista un $\bar{\alpha} \in [0, 1]$ tale che la distribuzione della variabile aleatoria

$$X^{\bar{\alpha}}$$

sia proprio \mathbb{P}_0 . Lo scopo di questo paragrafo è quello di determinare una successione di stimatori fortemente consistente di $\bar{\alpha}$.

Definiamo la funzione

$$G : [0, 1] \times \mathbb{R} \rightarrow [0, 1]$$

$$(\alpha, t) \mapsto \mathbb{P}\{X^\alpha \leq t\},$$

che ad ogni coppia (α, t) associa il valore assunto dalla funzione di ripartizione di X^α in t .

Indichiamo, per $n \in \mathbb{N}$, con \hat{G}_n la funzione

$$[0, 1] \times \Omega \times \mathbb{R} \rightarrow [0, 1]$$

$$(\alpha, \omega, t) \mapsto \sum_{i=0}^n I\{X_i^\alpha \leq t\},$$

che associa alla terna (ω, α, t) il valore assunto dalla funzione di ripartizione empirica di X_i^α in t .

Proposizione 4.3.1. *Esiste un insieme A con $\mathbb{P}\{A\} = 0$ tale che, se $\omega \notin A$, per $n \rightarrow \infty$*

$$\sup_{t, \alpha} |\hat{G}_n(\alpha, \omega, t) - G(\alpha, t)| \rightarrow 0.$$

Dimostrazione. Fissato $\alpha_1 \in [0, 1]$, il teorema di Glivenko Cantelli assicura che esiste un insieme $A \in \mathcal{F}$ con $\mathbb{P}\{A\} = 0$ tale che, per $n \rightarrow \infty$

$$d(\hat{G}_n(\alpha_1, \omega, t), G(\alpha_1, t)) \rightarrow 0$$

se $\omega \notin A$. Scegliamo $\alpha_2 \in [0, 1]$, allora si ha

$$X_i^{\alpha_1} \leq t \text{ se e solo se } X_i^{\alpha_2} \leq t^{\frac{\alpha_2}{\alpha_1}}.$$

Dunque, dato $\omega \in \Omega$

$$|\hat{G}_n(\alpha_1, \omega, t) - G(\alpha_1, t)| = |(\hat{G}_n(\alpha_2, \omega, t^{\frac{\alpha_2}{\alpha_1}}) - G(\alpha_2, t^{\frac{\alpha_2}{\alpha_1}}))|.$$

Passando all'estremo superiore su $t \in \mathbb{R}$ otteniamo che:

$$\sup_{t \in \mathbb{R}} |\hat{G}_n(\alpha_2, \omega, t) - G(\alpha_2, t)| = \sup_{t \in \mathbb{R}} |\hat{G}_n(\alpha_1, \omega, t) - G(\alpha_1, t)|,$$

quindi se $\omega \notin A$, per $n \rightarrow \infty$

$$\sup_{t, \beta} |\hat{G}_n(\beta, \omega, t) - G(\beta, t)| \rightarrow 0.$$

□

Consideriamo il campione di variabili trasformate $X_1^\alpha, \dots, X_n^\alpha$ e sia $p_n(\alpha, \omega)$ la variabile aleatoria *p-value* del test di Kolmogorov Smirnov di ipotesi nulla $H_0 : G(\alpha, t) = F_0$. Dunque:

$$p_n(\alpha, \omega) = 1 - F^*(\sqrt{n} \cdot d(\hat{G}_n(\alpha, \omega, t), F_0(t))).$$

Proposizione 4.3.2. *Per ogni $n \in \mathbb{N}$ esiste un insieme A_n di probabilità nulla tale che, se $\omega \notin A_n$, la funzione*

$$\alpha \mapsto p_n(\alpha, \omega)$$

è continua.

Dimostrazione. F^* è continua: ci basta quindi dimostrare che, quasi certamente, la funzione

$$\alpha \mapsto d(\hat{G}_n(\alpha, \omega, t), F_0(t)),$$

è una funzione continua. A meno di riordinare le osservazioni, possiamo supporre che $x_1 = X_1(\omega) < \dots < x_n = X_n(\omega)$ (in realtà così supponiamo anche che le osservazioni siano distinte: il caso in cui ve ne siano di coincidenti non porta nessuna differenza nella dimostrazione, ma ne complica molto la scrittura).

Per definizione di continuità, scelto $\alpha_0 \in [0, 1]$, fissato $\epsilon > 0$ dobbiamo determinare $c_\epsilon > 0$ tale che se $\alpha \in [0, 1]$ è tale che $|\alpha - \alpha_0| < c_\epsilon$, si ha

$$\left| \sup_t |\hat{G}_n(\alpha, \omega, t) - F_0(t)| - \sup_t |\hat{G}_n(\alpha_0, \omega, t) - F_0(t)| \right| < \epsilon.$$

Dividiamo in tre passi la dimostrazione, per renderla più comprensibile.

Passo 1. F_0 è, per ipotesi, continua, quindi è uniformemente continua sul compatto

$$[x_1^{\alpha_0} - 1, x_n^{\alpha_0} + 1].$$

Esiste dunque $d_\epsilon > 0$ tale che, se $x, y \in [x_1^{\alpha_0} - 1, x_n^{\alpha_0} + 1]$ e $|x - y| \leq d_\epsilon$

$$|F_0(x) - F_0(y)| < \epsilon.$$

Possiamo supporre, a meno di scegliere nuovamente d_ϵ , che esso sia minore di 1 e che gli intervalli simmetrici di centro x_i e raggio d_ϵ siano disgiunti.

Passo 2. La funzione

$$\alpha \mapsto (x_1^\alpha, \dots, x_n^\alpha)$$

è continua, dunque esiste c_ϵ tale che, se $\alpha \in [0, 1]$ e $|\alpha - \alpha_0| \leq c_\epsilon$

$$\|(x_1^\alpha - x_1^{\alpha_0}, \dots, x_n^\alpha - x_n^{\alpha_0})\| < d_\epsilon.$$

Passo 3. Consideriamo $t \in \mathbb{R}$, α tale che $|\alpha - \alpha_0| < c_\epsilon$ e l'insieme

$$U = \cup_{i=1}^n (x_i^{\alpha_0} - d_\epsilon, x_i^{\alpha_0} + d_\epsilon).$$

Fissato i , la funzione

$$t \mapsto I\{x_i^{\alpha_0} \leq t\}$$

è una funzione costante a tratti con un unico punto di discontinuità in $x_i^{\alpha_0}$.

Per il passo 2 si ha $|x_i^\alpha - x_i^{\alpha_0}| < d_\epsilon$, quindi se $t \notin U$

$$I\{x_i^{\alpha_0} \leq t\} = I\{x_i^\alpha \leq t\}.$$

Supponiamo che $t \in U$. U è, per il passo 1, unione disgiunta di interni: esiste un unico x_i tale che $|t - x_i^{\alpha_0}| < d_\epsilon$. In questo caso potrebbe essere

$$I\{x_i^{\alpha_0} \leq t\} \neq I\{x_i^\alpha \leq t\}.$$

ma, poiché $|x_i^\alpha - x_i^{\alpha_0}| < d_\epsilon$, esiste $t' \in \mathbb{R}$ tale che $|t - t'| < d_\epsilon$ e

$$\{x_i^{\alpha_0} \leq t\} = I\{x_i^\alpha \leq t'\}.$$

Quindi, riassumendo i due casi, per ogni $t \in \mathbb{R}$ esiste $t' \in \mathbb{R}$ tale che $|t-t'| < d_\epsilon$ e

$$\hat{G}_n(\alpha_0, \omega, t) = \hat{G}_n(\alpha, \omega, t').$$

Quindi n

$$\begin{aligned} |\hat{G}_n(\alpha_0, \omega, t) - F_0(t)| &= |(\hat{G}_n(\alpha, \omega, t') - F_0(t')) + (F_0(t') - F_0(t))| \leq \\ &\leq |(\hat{G}_n(\alpha, \omega, t') - F_0(t'))| + |F_0(t') - F_0(t)|. \end{aligned}$$

Per il punto 1 abbiamo $|F_0(t') - F_0(t)| \leq \epsilon$, dunque

$$|\hat{G}_n(\alpha, \omega, t') - F_0(t')| + |F_0(t') - F_0(t)| \leq |(\hat{G}_n(\alpha, \omega, t') - F_0(t'))| + \epsilon$$

per cui, passando agli estremi superiori

$$\sup_{t \in \mathbb{R}} |(\hat{G}_n(\alpha_0, \omega, t) - F_0(t))| - \sup_{t \in \mathbb{R}} |(\hat{G}_n(\alpha, \omega, t) - F_0(t))| \leq \epsilon$$

Invertendo i ruoli di α e α_0 otteniamo anche la disuguaglianza

$$\sup_{t \in \mathbb{R}} |(\hat{G}_n(\alpha_0, \omega, t) - F_0(t))| - \sup_{t \in \mathbb{R}} |(\hat{G}_n(\alpha, \omega, t) - F_0(t))| \geq -\epsilon,$$

da cui

$$\left| \sup_t |\hat{G}_n(\alpha, \omega, t) - F_0(t)| - \sup_t |\hat{G}_n(\alpha_0, \omega, t) - F_0(t)| \right| < \epsilon.$$

□

Fissati $n \in \mathbb{N}$ e $\omega \notin A_n$, la funzione

$$\alpha \mapsto p_n(\alpha, \omega)$$

è quindi continua su $[0, 1]$, che è compatto, perciò esiste un massimo globale ed un relativo punto di massimo. Intendiamo scegliere questo punto di massimo come stimatore di $\bar{\alpha}$, bisogna tuttavia formalizzare con cautela questa scelta, perché il punto di massimo potrebbe non essere unico e perché uno stimatore deve essere una funzione misurabile.

Definiamo la funzione $p_n^{-1}(\omega, x)$ in questo modo:

$$p_n^{-1} : \Omega \times [0, 1] \rightarrow [0, 1]$$

$$(\omega, x) \mapsto \begin{cases} \inf\{\alpha | p_n(\alpha, \omega) \geq x\} & \text{se } \{\alpha | p_n(\alpha, \omega) \geq x\} \neq \emptyset \\ -1 & \text{se } \{\alpha | p_n(\alpha, \omega) \geq x\} = \emptyset. \end{cases}$$

Poniamo quindi

$$\hat{\alpha}_n(\omega) = \begin{cases} p_n^{-1}(\omega, \sup_{\alpha \in [0,1]} p_n(\alpha, \omega)) & \text{se } \omega \notin A_n \\ -1 & \text{se } \omega \in A_n. \end{cases}$$

È dunque ben definita la successione di funzioni $\{\hat{\alpha}_n(\omega)\}_{n \in \mathbb{N}}$: a meno dell'insieme $\cup_{n \in \mathbb{N}} A_n$ queste funzioni associano a ω un punto di massimo globale della funzione $p_n(\alpha, \omega)$; nel lemma che segue dimostriamo che si tratta di una successione di variabili aleatorie.

Lemma 4.3.1. *Per ogni $n \in \mathbb{N}$, la funzione*

$$\omega \mapsto \hat{\alpha}_n(\omega)$$

è una variabile aleatoria.

Dimostrazione. Consideriamo solo la definizione su A_n^c , visto che sull'insieme misurabile A_n la funzione $\hat{\alpha}_n$ è costante. Se $\omega \notin A_n$, la funzione

$$\alpha \rightarrow p_n(\alpha, \omega)$$

è continua per la proposizione precedente, dunque

$$\sup_{\alpha \in [0,1]} p_n(\alpha, \omega) = \sup_{\alpha \in \mathbb{Q} \cap [0,1]} p_n(\alpha, \omega).$$

Dal momento che per ogni $\alpha \in [0, 1]$ la funzione

$$\omega \mapsto p_n(\alpha, \omega)$$

è misurabile, anche

$$\omega \mapsto \sup_{\alpha \in [0,1]} p_n(\alpha, \omega)$$

è una funzione misurabile. Per la stessa ragione, fissato $x \in [0, 1]$, la funzione $p_n^{-1}(x, \alpha)$ è misurabile, dunque

$$p_n^{-1} : \Omega \times [0, 1] \rightarrow [0, 1]$$

$$(\omega, x) \mapsto \begin{cases} \inf\{\alpha | p_n(\alpha, \omega) \geq x\} & \text{se } \{\alpha | p_n(\alpha, \omega) \geq x\} \neq \emptyset \\ -1 & \text{se } \{\alpha | p_n(\alpha, \omega) \geq x\} = \emptyset. \end{cases}$$

è una funzione misurabile. □

Vogliamo dimostrare che $\{\hat{\alpha}_n\}_{n \in \mathbb{N}}$ è una successione di stimatori fortemente consistente di $\bar{\alpha}$. Premettiamo la

Proposizione 4.3.3. *Supponiamo che la successione di variabili aleatorie definite su $(\Omega, \mathcal{F}, \mathbb{P})$*

$$\{p_n(\bar{\alpha}, \omega)\}_{n \in \mathbb{N}}$$

converga puntualmente. Allora si ha, per \mathbb{P} -quasi ogni ω

$$p(\alpha, \omega) = \lim_n p_n(\alpha, \omega) = \begin{cases} 0 & \text{se } \alpha \neq \bar{\alpha} \\ c(\omega) \neq 0 & \text{se } \alpha = \bar{\alpha}. \end{cases}$$

Dimostrazione. Se $\alpha \neq \bar{\alpha}$ la funzione di ripartizione di X^α è diversa da F_0 . Per la proposizione 4.3.1 esiste un insieme A con $\mathbb{P}\{A\} = 0$ tale che se $\omega \notin A$, per ogni $\alpha \in [0, 1]$

$$\lim_n d(\hat{G}_n(\alpha, \omega, t), F_0) \neq 0,$$

dunque se $\omega \notin A$:

$$\lim_n p_n(\alpha, \omega) = 0.$$

Ora consideriamo il caso $\alpha = \bar{\alpha}$. In questo caso

$$p_n(\bar{\alpha}, \omega) \xrightarrow{\mathcal{L}} U$$

dove Y ha distribuzione uniforme su $[0, 1]$. Per ipotesi esiste anche il limite puntuale

$$p(\bar{\alpha}, \omega) = \lim_n p_n(\bar{\alpha}, \omega),$$

che dunque avrà distribuzione uniforme su $[0, 1]$. Siccome

$$\mathbb{P}\{p(\bar{\alpha}, \omega) = 0\} = 0,$$

\mathbb{P} -quasi certamente

$$p(\alpha, \omega) = \lim_n p_n(\alpha, \omega) = \begin{cases} 0 & \text{se } \alpha \neq \bar{\alpha} \\ c(\omega) \neq 0 & \text{se } \alpha = \bar{\alpha}. \end{cases}$$

□

Per la proposizione 4.3.2 le funzioni $p_n(\alpha, \omega)$ sono, almeno quasi certamente, continue, mentre il loro limite su n non lo è: non possiamo sperare che, a ω fissato, la convergenza sia uniforme. Vogliamo però dimostrare che esiste un determinato sottoinsieme di $[0, 1]$ su cui ciò avviene. Abbiamo bisogno del seguente lemma.

Lemma 4.3.2. *Sia Y una variabile aleatoria reale quasi certamente positiva, con funzione di ripartizione F . Sia F_α la funzione di ripartizione della variabile aleatoria Y^α . La funzione*

$$\alpha \mapsto \sup_t |F_\alpha(t) - F(t)| = d(F_\alpha, F)$$

è una funzione crescente su $[1, \infty)$ e decrescente su $[0, 1)$.

Dimostrazione. Dimostriamo solo che la funzione è crescente su $[1, \infty)$, l'altra parte è del tutto analoga. Poiché

$$|F_\alpha(t) - F(t)| = |\mathbb{P}\{Y \leq t^{\frac{1}{\alpha}}\} - \mathbb{P}\{Y \leq t\}|,$$

distinguiamo due casi: se $t \geq 1$

$$|\mathbb{P}\{Y \leq t^{\frac{1}{\alpha}}\} - \mathbb{P}\{Y \leq t\}| = \mathbb{P}\{t^{\frac{1}{\alpha}} \leq Y \leq t\},$$

mentre se $t < 1$

$$|\mathbb{P}\{Y \leq t^{\frac{1}{\alpha}}\} - \mathbb{P}\{Y \leq t\}| = \mathbb{P}\{t \leq Y \leq t^{\frac{1}{\alpha}}\}.$$

Entrambe sono funzioni crescenti di α , quindi se $\alpha_2 > \alpha_1 > 1$

$$d(F_{\alpha_1}, F) = \sup_t |F_{\alpha_1}(t) - F(t)| \leq \sup_t |F_{\alpha_2}(t) - F(t)| = d(F_{\alpha_2}, F).$$

□

Per dimostrare che $\{\hat{\alpha}_n\}_{n \in \mathbb{N}}$ è una successione di stimatori fortemente consistente di $\bar{\alpha}$ abbiamo bisogno di rafforzare il risultato sulla convergenza puntuale della variabile $p_n(\alpha, \omega)$.

Lemma 4.3.3. *Nelle stesse ipotesi della proposizione 4.3.3, esiste un insieme A con $\mathbb{P}\{A\} = 0$ tale che se $\omega \notin A$ per ogni*

$$0 < \epsilon < \min\{\bar{\alpha}, 1 - \bar{\alpha}\}.$$

la convergenza delle funzioni

$$p_n(\alpha, \omega) \rightarrow 0$$

sull'insieme

$$[0, \alpha_1] \cup [\alpha_2, 1]$$

è uniforme, dove $\alpha_1 = \bar{\alpha} - \epsilon$ e $\alpha_2 = \bar{\alpha} + \epsilon$.

Dimostrazione. Poiché

$$p_n(\alpha, \omega) = 1 - F^*(\sqrt{n}d(\hat{G}_n(\alpha, \omega, t), F_0(t)))$$

e F^* è continua, condizione sufficiente per la convergenza uniforme alla costante 0 sull'insieme considerato è che la convergenza

$$\sqrt{n}d(\hat{G}_n(\alpha, \omega, t), F_0(t)) \rightarrow \infty$$

sia uniforme. Ancora, visto che $\sqrt{n} \rightarrow \infty$, ci basta dimostrare che la successione di funzioni

$$\alpha \mapsto \sup_t |\hat{G}_n(\alpha, \omega, t) - F_0(t)|$$

è definitivamente maggiore di un $\delta > 0$; cioè che esiste $\delta > 0$ e $n_\delta > 0$ tale che per ogni $n > n_\delta$, se

$$\begin{aligned} \alpha &\in [0, \alpha_1] \cup [\alpha_2, 1], \\ \sup_t |\hat{G}_n(\alpha, \omega, t) - F_0(t)| &> \delta. \end{aligned}$$

Scegliamo $\delta > 0$ in modo che

$$0 < 2\delta < \min\{d(G(\alpha_1, t), F_0(t)), d(G(\alpha_2, t), F_0(t))\}.$$

Se $\alpha < \alpha_1$ o $\alpha > \alpha_2$, il lemma 4.3.2 garantisce che

$$\sup_t |G(\alpha, t) - F_0(t)| > 2\delta.$$

Per la proposizione 4.3.1, esiste A di probabilità nulla (che non dipende da ϵ) tale che se $\omega \notin A$ la convergenza

$$\hat{G}_n(\alpha, \omega, t) \rightarrow G(\alpha, t)$$

è uniforme in α e in t , quindi esiste $n_\delta \in \mathbb{N}$ tale che, per $n > n_\delta$ per ogni $\alpha \in [0, 1]$

$$\sup_t |\hat{G}_n(\alpha, \omega, t) - G(\alpha, t)| < \delta.$$

Scelto $\alpha \in [0, \alpha_1] \cup [\alpha_2, 1]$ esiste $t \in \mathbb{R}^+$ tale che $|G(\alpha, \omega, t) - F_0(t)| > 2\delta$, per tale t si avrà:

$$|\hat{G}_n(\alpha, \omega, t) - F_0(t)| \geq |G(\alpha, t) - F_0(t)| - |\hat{G}_n(\alpha, \omega, t) - G(\alpha, t)| > \delta,$$

come si voleva dimostrare. \square

L'ultimo risultato che dimostriamo è la forte convergenza della successione $\{\alpha_n\}_{n \in \mathbb{N}}$ di stimatori di $\bar{\alpha}$.

Proposizione 4.3.4. *La successione di variabili aleatorie $\hat{\alpha}_n$, nelle ipotesi enunciate in 4.3.3, è una successione di stimatori fortemente consistente di $\bar{\alpha}$.*

Dimostrazione. Sia A l'insieme di probabilità nulla la cui esistenza è stata dimostrata nel lemma 4.3.3. Supponiamo che esista un insieme $B \subset A^c$ di probabilità non nulla su cui il limite di $\hat{\alpha}_n$ non esiste, oppure è diverso da $\bar{\alpha}$. Sia $\omega \in B$: possiamo estrarre una sottosuccessione

$$\hat{\alpha}_{n_k}(\omega)$$

convergente a $\beta \neq \bar{\alpha}$, con $\beta \in [0, 1]$. Applicando il lemma 4.3.3 con $\epsilon < |\beta - \bar{\alpha}|$: anche $p_{n_k}(\alpha_{n_k}, \omega) \rightarrow 0$. Gli $\hat{\alpha}_n(\omega)$ sono stati scelti come punti di massimo della funzione

$$\alpha \rightarrow p_n(\alpha),$$

mentre, per il lemma 4.3.3

$$p_n(\bar{\alpha}, \omega) \rightarrow c(\omega) \neq 0.$$

Dunque per ogni $\omega \in B$ esiste $n = n(\omega) \in \mathbb{N}$ per cui

$$p_n(\bar{\alpha}, \omega) > p_n(\hat{\alpha}_n, \omega) \text{ e } \bar{\alpha} \neq \hat{\alpha}_n$$

contro l'ipotesi di massimalità degli $\hat{\alpha}_n$. □

4.3.3 Risultati ottenuti e confronto con il metodo di Box Cox.

Il metodo esposto nel paragrafo precedente è stato costruito sul test di Kolmogorov-Smirnov nel caso di ipotesi nulla semplice, con F_0 funzione di ripartizione di una probabilità concentrata su \mathbb{R}^+ . Per applicarlo, dunque, proseguiamo in questo modo. Appliciamo al primo campione il metodo di Box Cox e, confidando nella bontà della stima trovata, stimiamo media $\hat{\mu}$ e varianza $\hat{\sigma}^2$ della relativa distribuzione normale con gli stimatori di massima verosimiglianza utilizzando i dati trasformati del primo campione (cioè i dati i relativi alla prima visita). Appliciamo poi ai campioni successivi il nostro

metodo, con F distribuzione gaussiana troncata all'insieme $(0, \infty]$ con media \hat{m} e varianza $\hat{\sigma}^2$.

Anche in questo caso calcoliamo una stima di $\bar{\alpha}$ per ognuno dei 44 campioni rimanenti. Il punto di massimo della funzione $p_n(\alpha, \omega)$ è stato stimato discretizzando il dominio della funzione all'insieme:

$$\left\{ \frac{n}{100} \mid n = 1, \dots, 100 \right\}.$$

Nella seguente tabella riportiamo i valori ottenuti per i campioni 2-6:

Tabella 4.4: Stime di $\bar{\alpha}$ relative a diversi campioni.

	Visita 1	Visita 2	Visita 3	Visita 4
$\hat{\alpha}$	0.25	0.33	0.36	0.18

Siccome nel paragrafo precedente abbiamo dimostrato che $\{\hat{\alpha}_n\}$ è una successione di stimatori fortemente consistente del parametro $\bar{\alpha}$ cercato, tale sarà anche la media aritmetica fra gli stimatori relativi a campioni diversi. In questo modo otteniamo la stima complessiva:

$$\hat{\alpha} = 0.22.$$

Le trasformazioni parametrizzate da α in questo metodo sono diverse da quelle utilizzate in Box Cox, ma possono esservi riportate mediante una funzione affine. Dal momento che, se X è una variabile aleatoria gaussiana, lo è anche $aX + b$ (con $a \neq 0$, se non si considerano casi degeneri), possiamo confrontare direttamente le stime ottenuti coi due metodi: per 36 delle 44 visite, il valore stimato con l'ultimo metodo è interno all'intervallo di confidenza di livello 0.05 stimato attraverso Box Cox. Nel nostro caso, perciò, i due metodi sembrano condurre a risultati compatibili: d'ora in poi supporremo che il processo

$$\{f_\lambda(X(t))\}_{t \in \mathbb{R}^+} \quad \text{con } \lambda = 0.2$$

sia un processo gaussiano.

Capitolo 5

Inferenza statistica sui parametri degli individui sani

Supponiamo che $\{X(t)\}$ sia il processo che modella il decadimento del numero di linfociti $CD4+$ per mm^3 di sangue in una persona sieronegativa. Nel capitolo precedente abbiamo modificato le ipotesi fatte in [7] e abbiamo determinato una trasformazione f_λ , alternativa al logaritmo, per cui

$$\{f_\lambda(X(t))\}_{t \in \mathbb{R}^+}$$

è un processo gaussiano. In questo capitolo vogliamo stimare la media e la funzione di covarianza di questo processo, che, come abbiamo visto nel capitolo 4, ne caratterizzano le proprietà statistiche.

Abbiamo a disposizione 45 misurazioni ripetute per ognuno dei 2548 soggetti. Il fatto di avere più visite per ciascun individuo consente una stima della funzione di covarianza più approfondita rispetto a quella di [7]. Tale stima sarà utile per testare la validità del modello per i sieropositivi, nei capitoli 7 e 8.

La derivazione degli stimatori della media e di alcuni valori della funzione di covarianza in questo capitolo si basano sostanzialmente sulla legge dei grandi numeri: ricaviamo uno stimatore corretto basato sui dati di un singolo paziente, per poi considerare medie campionarie di questi stimatori e dare dei risultati asintotici sul numero n dei soggetti considerati. Abbiamo considerato anche l'inferenza parametrica sulla funzione di covarianza sotto l'ipotesi che il processo sia di Markov.

In questo capitolo discutiamo anche le distribuzioni esatte degli stimatori e le motivazioni che ci hanno portato a sceglierli. Un paragrafo è dedicato

alle distribuzioni di forme quadratiche di vettori gaussiani, che utilizzeremo anche in seguito.

5.1 Notazione utilizzata

In [7] gli autori propongono di modellizzare la concentrazione di linfociti $CD4+$ per mm^3 di sangue in un individuo sieronegativo con un processo stocastico stazionario log-gaussiano

$$\{X(t)\}_{t \in \mathbb{R}^+}.$$

Abbiamo discusso questa ipotesi: noi supporremo che

$$\{f_\lambda(X(t))\}_{t \in \mathbb{R}^+}$$

sia un processo gaussiano, dove f_λ è la trasformazione individuata nel corso del capitolo 4. Indichiamo con μ la media di questo processo e con Ξ la sua funzione di covarianza.

I dati a nostra disposizione riguardano misurazioni effettuate su 2548 individui, ripetute a intervalli di tempo regolari di sei mesi per 45 volte.

Possiamo identificare ciascun soggetto considerato con un numero naturale i , dove $1 \leq i \leq 2548$. Indichiamo il processo che modella la concentrazione di linfociti $CD4+$ per mm^3 di sangue nell' i -esimo individuo con

$$\{X_i(t)\}_{t \in \mathbb{R}^+}.$$

Supponiamo che $t = 0$ coincida con l'inizio del monitoraggio (ciò è lecito perché il processo è stazionario). Le 45 misurazioni dell' i -esimo individuo sono le osservazioni delle prime 45 variabili del processo a tempi discreti.

$$\{X_i(nh)\}_{n \in \mathbb{N}},$$

dove h è l'intervallo di tempo trascorso fra una visita e la successiva. Indicheremo con $Y_i(n)$ il processo gaussiano

$$\{Y_i(n)\}_{n \in \mathbb{N}} = \{f_\lambda(X_i(nh))\}_{n \in \mathbb{N}}.$$

Vista la definizione di questo processo, secondo la modifica fatta al modello proposto in [7], funzione di covarianza e media di $\{Y_i(n)\}_{n \in \mathbb{N}}$ sono date da

$$\mathbb{E}\{Y_i(n)\} = \mu \quad \text{e} \quad \Gamma(q) = \text{cov}\{Y_i(n)Y_i(n+q)\} = \Xi(qh).$$

Non tutti gli individui si sono presentati ad ogni visita: indicheremo perciò con m_i il numero di osservazioni disponibili per l' i -esimo individuo. Supporremo poi che i processi relativi a pazienti diversi siano indipendenti, dunque che $\{Y_i(n)\}_{n \in \mathbb{N}}$ sia indipendente da $\{Y_j(n)\}_{n \in \mathbb{N}}$ se $i \neq j$. Infine, allo scopo di provare i risultati asintotici, supporremo di avere una successione $\{Y_i\}_{i \in \mathbb{N}}$ di processi con le caratteristiche appena descritte.

5.2 La stima di μ

In questo capitolo vogliamo ricavare e studiare le proprietà di alcuni stimatori di μ . Ricordiamo che confronteremo sempre gli stimatori tramite il costo quadratico medio.

5.2.1 Determinazione degli stimatori di μ

Il primo problema che vogliamo affrontare è l'inferenza statistica su μ a partire dai dati relativi all' i -esimo paziente.

Consideriamo il processo gaussiano stazionario $\{Y_i(n)\}_{n \in \mathbb{N}}$, definito sullo spazio $(\Omega, \mathcal{F}, \mathbb{P})$ e il modello statistico indotto dal campione

$$(Y_i(0), \dots, Y_i(m_i - 1)).$$

Esso è un vettore gaussiano, con vettore delle medie

$$(\mu, \dots, \mu).$$

Vogliamo determinare uno stimatore di μ : supponiamo, per semplicità, di volerlo cercare lineare. Nel paragrafo 5.2.2 si dimostra che esiste uno stimatore ottimale fra quelli lineari e corretti di μ . Le ragioni che ci inducono a non considerare tale stimatore sono descritte più avanti: il motivo principale è comunque il fatto che questo stimatore necessita la conoscenza della matrice di covarianza del vettore aleatorio $(Y(1), \dots, Y(m_i - 1))$ e dunque di alcuni valori della funzione di covarianza Γ , su cui non abbiamo alcuna informazione. Consideriamo allora lo stimatore dei minimi quadrati:

$$\bar{Y}_i = \frac{1}{m_i} \sum_{k=0}^{m_i-1} Y_i(k), \quad (5.1)$$

che è uno stimatore corretto di μ . Grazie alle ipotesi fatte sul processo, possiamo descrivere la distribuzione degli stimatori. Per definizione di vettore gaussiano, \bar{Y}_i è una variabile gaussiana, con media μ . Calcoliamone la varianza:

$$\begin{aligned}\text{Var}\{\bar{Y}_i\} &= \frac{1}{m_i^2} \sum_{k=0}^{m_i-1} \sum_{j=0}^{m_i-1} \text{cov}\{X_k, X_j\} = \\ &= \frac{1}{m_i^2} \sum_{|h| \leq m_i-1} (m_i - |h|) \Gamma(h) = \frac{1}{m_i} \sum_{|h| \leq m_i-1} \left(1 - \frac{|h|}{m_i}\right) \Gamma(h),\end{aligned}$$

quindi

$$\bar{Y}_i \sim N\left(\mu, \frac{1}{m_i} \sum_{|h| \leq m_i-1} \left(1 - \frac{|h|}{m_i}\right) \Gamma(h)\right).$$

Adesso consideriamo le misure relative ad n individui: consideriamo il modello statistico indotto dalle statistiche $\bar{Y}_1, \dots, \bar{Y}_n$. Cerchiamo uno stimatore per μ : si ripropone lo stesso problema del paragrafo precedente: uno stimatore lineare ottimale esiste, ma presuppone la conoscenza della covarianza del vettore

$$(\bar{Y}_1, \dots, \bar{Y}_n)$$

e dunque della funzione di covarianza Γ . Anche in questo caso scegliamo lo stimatore dei minimi quadrati:

$$C_n = n^{-1} \sum_{i=1}^n \bar{Y}_i.$$

C_n è uno stimatore corretto di μ . Per derivare le proprietà asintotiche della successione di stimatori

$$\{C_n\}_{n \in \mathbb{N}}$$

è necessaria la seguente versione del teorema forte dei grandi numeri, dovuta a Kolmogorov (per una dimostrazione vedi [26]).

Teorema 5.2.1. *Sia $\{X_i\}_{i \in \mathbb{N}}$ una successione di variabili aleatorie indipendenti in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, con uguale valore atteso, tale che*

$$\sum_{i=1}^{\infty} \frac{\text{Var}\{X_i\}}{i^2} < \infty,$$

allora la successione di variabili aleatorie

$$\frac{1}{n} \sum_{i=1}^n X_i$$

converge puntualmente \mathbb{P} -quasi certamente alla costante $\mathbb{E}\{X_i\}$.

Proposizione 5.2.1. *La successione di variabili aleatorie*

$$C_n = \frac{1}{n} \sum_{i=1}^n \bar{Y}_i$$

è una successione di stimatori fortemente consistente di μ .

Dimostrazione. Per il teorema 5.2.1 ci basta verificare che

$$\sum_{i=1}^{\infty} \frac{\text{Var}\{\bar{Y}_i\}}{i^2} < \infty.$$

La funzione di covarianza Γ è definita positiva, quindi, per quanto detto nel capitolo 2

$$|\Gamma(s)| \leq \Gamma(0) \quad \text{per ogni } s \in \mathbb{R},$$

di conseguenza

$$\begin{aligned} \text{Var}\bar{Y}_i &= \frac{1}{m_i} \sum_{|h| \leq m_i - 1} \left(1 - \frac{|h|}{m_i}\right) \Gamma(h) \leq \\ & \frac{1}{m_i} \sum_{|h| \leq m_i - 1} \left(1 - \frac{|h|}{m_i}\right) |\Gamma(h)| \leq \Gamma(0). \end{aligned}$$

Quindi

$$\sum_i \frac{\text{Var}\{\bar{Y}_i\}}{i^2} \leq \Gamma(0) \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty,$$

come volevamo dimostrare. □

Nel caso disponessimo di una sola misura per paziente, lo stimatore C_n , basato sulle variabili aleatorie

$$Y_1(0), \dots, Y_n(0)$$

coincide evidentemente con lo stimatore di massima verosimiglianza del campione $Y_1(0), \dots, Y_n(0)$, che è a variabili indipendenti ed identicamente distribuite con legge gaussiana. Quest'ultimo è il metodo che viene suggerito in [7] per la stima di μ . Con la prossima proposizione, dimostriamo che vi è un effettivo guadagno nel considerare più di una misura per singolo paziente.

Proposizione 5.2.2. *Supponiamo che la successione $\{m_i\}_{i \in \mathbb{N}}$ sia tale che $m_i \geq 2$ per ogni $i \in \mathbb{N}$. Consideriamo la successione di stimatori*

$$C_n = \frac{1}{n} \sum_{i=1}^n \bar{Y}_i,$$

dove gli \bar{Y}_i sono definiti da (5.1) e la successione di stimatori

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n Y_i(0).$$

Allora per ogni $n \in \mathbb{N}$

$$\text{Var}\{C_n\} \leq \frac{1}{2} \text{Var}\{\hat{\mu}_n\}.$$

Dimostrazione. Si tratta di un semplice calcolo. Notiamo dapprima che se $k < n \in \mathbb{N}$

$$\sum_{|h| < n} (n - h) = \binom{n}{2}.$$

Ricordando inoltre che $|\Gamma(s)| \leq |\Gamma(0)|$ per ogni $s \in \mathbb{R}$:

$$\begin{aligned} \text{Var}\bar{Y}_i &\leq \frac{1}{m_i} \sum_{|h| \leq m_i - 1} \left(1 - \frac{|h|}{m_i}\right) \Gamma(0) \leq \frac{1}{m_i^2} \binom{m_i}{2} \Gamma(0) = \\ &= \frac{(m_i)(m_i - 1) \Gamma(0)}{m_i^2} \frac{1}{2}. \end{aligned}$$

Quindi

$$\text{Var}\{C_n\} = \frac{1}{n^2} \sum_i^n \text{Var}\{\bar{Y}_i\} = \frac{\Gamma(0)}{2n^2} \sum_{i=1}^n \frac{m_i(m_i - 1)}{m_i^2} \leq \frac{\Gamma(0)}{2n} = \frac{\text{Var}\{\hat{\mu}_n\}}{2}.$$

□

5.2.2 È possibile la scelta di uno stimatore migliore per μ ?

Lo scopo di questo paragrafo è chiarire ciò che abbiamo detto nella sezione precedente: in due diverse situazioni sarebbe stato possibile scegliere uno stimatore preferibile a quello che abbiamo effettivamente usato. Qui tentiamo di motivare questa scelta.

Supponiamo che

$$Y = (Y_1, \dots, Y_n)^t$$

sia un vettore aleatorio definito su $(\Omega, \mathcal{F}, \mathbb{P})$, con vettore delle medie $(\mu, \dots, \mu)^t$ e matrice di covarianza Q . Definite le variabili aleatorie

$$\epsilon_i = Y_i - \mu,$$

possiamo scrivere

$$(Y_1, \dots, Y_n)^t = \mathbf{1}^t \mu + (\epsilon_1, \dots, \epsilon_n)^t$$

dove $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ è un vettore centrato con matrice di covarianza Q e

$$\mathbf{1} = (1, \dots, 1)^t.$$

Il problema dell'inferenza statistica su μ è dunque un problema di regressione lineare. In questa situazione esistono degli stimatori con proprietà ottimali per μ .

Definizione 5.2.1. Sia $(\Omega, \mathcal{F}, (\mathbb{P}^\theta)_{\theta \in \Theta})$ un modello statistico indotto dal campione

$$(Y_1, \dots, Y_n)$$

e sia $f : \Theta \rightarrow \mathbb{R}^k$ una funzione misurabile.

Una variabile aleatoria

$$X : \Omega \rightarrow \mathbb{R}^k$$

si dice uno stimatore lineare di $f(\theta)$ se è una combinazione lineare delle variabili Y_1, \dots, Y_n . Lo stimatore X si dice BLUE (Best Linear Unbiased Estimator) di $f(\theta)$ se è uno stimatore ottimale nella classe degli stimatori lineari e corretti di $f(\theta)$.

La proposizione che segue è un caso particolare del teorema di Gauss-Markov-Aitken, che sarà dimostrato nel capitolo 7. Essa garantisce che, nella situazione che abbiamo considerato, esiste uno stimatore BLUE di μ .

Proposizione 5.2.3. *Supponiamo che $(Y_1, \dots, Y_n)^t$ sia un vettore aleatorio definito su $(\Omega, \mathcal{F}, \mathbb{P})$ e che sia soddisfatto il modello di regressione*

$$(Y_1, \dots, Y_n)^t = \mu \mathbf{1} + (\epsilon_1, \dots, \epsilon_n)^t, \quad (5.2)$$

dove $\mu \in \mathbb{R}$ e il vettore aleatorio di errore $(\epsilon_1, \dots, \epsilon_n)^t$ è centrato con matrice di covarianza Q non singolare. Lo stimatore:

$$(\mathbf{1}^t Q^{-1} \mathbf{1})^{-1} \mathbf{1} Q^{-1} Y.$$

è uno stimatore BLUE di μ .

Questo stimatore ha il grosso inconveniente di coinvolgere la matrice di covarianza del vettore Y , quindi nel nostro caso la funzione di covarianza del processo, che è incognita. Supposto che essa sia non singolare, potremmo pensare di utilizzare lo stimatore di cui la proposizione 5.2.3 garantisce l'esistenza con una stima della matrice di covarianza, che può essere determinata con gli stimatori che verranno esposti nei paragrafi seguenti. Questo metodo, detto *feasible generalized least square* porta però a delle complicazioni, a cui accenniamo brevemente.

Da una parte Amemiya in [3], sotto alcune ipotesi particolari, ha provato che la successione degli stimatori *feasible* e quella dei BLUE convergono in legge ad una stessa variabile aleatoria; questo testimonierebbe a favore dell'adozione di questo tipo di stimatori, nel caso il campione sia di grossa taglia. D'altra parte, tuttavia, non c'è alcuna sicurezza che i nuovi stimatori abbiano un costo quadratico inferiore a quello degli stimatori dei minimi quadrati, specialmente quando i campioni sono di piccola taglia (si veda ad esempio [20], in cui sono presentati una serie di risultati di simulazioni Monte Carlo). Per queste ragioni, anziché procedere con lo stimatore *feasible generalized least square estimator* abbiamo sempre scelto la media campionaria.

5.3 Inferenza statistica sulla funzione di covarianza del processo

In questo paragrafo ci proponiamo di stimare $\Gamma(n)$ per $n \leq 5$. Queste stime saranno utili nel prosieguo della nostra analisi, specialmente nei capitoli 7 e 8. Per semplificare la trattazione matematica del problema, cerchiamo degli stimatori per la funzione

$$\Gamma^*(n) = \Gamma(n) + \mu^2,$$

che rappresenta la funzione che associa alle due variabili del processo relativo ad un qualsiasi individuo $Y_i(m)$ e $Y_i(m+n)$, il valore

$$\mathbb{E}\{Y_i(m)Y_i(m+n)\}.$$

Visto che nella sezione precedente abbiamo determinato una successione di stimatori di μ fortemente consistente, se riuscissimo a giungere a questo risultato anche per $\Gamma^*(n)$, otterremmo una successione di stimatori fortemente consistente di $\Gamma(n)$.

Nel prossimo paragrafo abbandoniamo, provvisoriamente, la notazione e le ipotesi introdotte in 5.1: esso è dedicato allo studio di particolari misure di probabilità su \mathbb{R} , che, oltre a consentire lo studio della distribuzione degli stimatori introdotti in 5.3.2, saranno utili nei capitoli successivi.

5.3.1 La distribuzione di forme quadratiche di vettori gaussiani

Definizione 5.3.1. *Supponiamo che X_1, \dots, X_n siano delle variabili aleatorie reali e indipendenti definite su $(\Omega, \mathcal{F}, \mathbb{P})$, tali che*

$$X_i \sim N(\mu_i, 1).$$

La distribuzione della variabile aleatoria

$$U = \sum_{i=1}^n X_i^2$$

si dice distribuzione χ^2 non centrale a n gradi di libertà con parametro di non centralità $\delta \geq 0$, dove

$$\delta^2 = \sum_{i=1}^n \mu_i^2.$$

Questa distribuzione è usualmente indicata col simbolo

$$\chi^2(n, \delta).$$

La proposizione seguente dimostra che la distribuzione della variabile aleatoria U dipende da μ_1, \dots, μ_n solo attraverso la quantità $\sum_{i=1}^n \mu_i^2$. Ciò giustifica la notazione $\chi^2(n, \delta)$.

Proposizione 5.3.1. *Supponiamo che X_1, \dots, X_n siano variabili aleatorie reali e indipendenti con $X_i \sim N(\mu_i, 1)$. Allora la distribuzione della variabile aleatoria*

$$\sum_{i=1}^n X_i^2$$

dipende solo da $\|(\mu_1, \dots, \mu_n)\| = \|\mu\|$.

Dimostrazione. Esiste un operatore ortogonale $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tale che

$$T(\mu_1, \dots, \mu_n) = (\|\mu\|, 0, \dots, 0).$$

Consideriamo il vettore aleatorio $(T(X_1), \dots, T(X_n))$. Esso è un vettore gaussiano, perché (X_1, \dots, X_n) è un vettore gaussiano e T è una trasformazione affine. Il vettore delle medie di $(T(X_1), \dots, T(X_n))$ è

$$T(\mu_1, \dots, \mu_n) = (\|\mu\|, 0, \dots, 0),$$

mentre la matrice di covarianza di $(T(X_1), \dots, T(X_n))$ è data, indicando con M_T la matrice associata all'operatore T , da

$$M_T \cdot I \cdot M_T^t = I.$$

Le variabili $T(X_1), \dots, T(X_n)$ sono dunque ancora indipendenti e $T(X_i)$ per $i \geq 2$ ha distribuzione gaussiana standard. Poiché T è un operatore ortogonale,

$$\sum_{i=1}^n X_i^2 = \sum_{i=1}^n T(X_i)^2 = T(X_1)^2 + \sum_{i=2}^n T(X_i)^2. \quad (5.3)$$

Per definizione

$$\sum_{i=2}^n T(X_i)^2 \sim \chi^2(n-1) \quad \text{e} \quad T(X_1) \sim N(\delta, 1);$$

le due variabili sono indipendenti, quindi $\|\mu\|$ e n identificano in maniera univoca la distribuzione $\chi^2(n, \delta)$. \square

Calcoliamo valore atteso e varianza delle distribuzioni $\chi^2(n, \delta)$.

Proposizione 5.3.2. *Sia X una variabile aleatoria reale, con*

$$X \sim \chi^2(n, \delta);$$

allora

$$\mathbb{E}\{X\} = \delta^2 + \mu \quad e \quad \text{Var}X = 2n + 4\delta^2.$$

Dimostrazione. Utilizziamo la decomposizione (5.3): la variabile X può essere scomposta nella somma

$$X = Y_1 + Y_2^2,$$

dove $Y_1 \sim \chi^2(n-1)$ e $Y_2 \sim N(\delta, 1)$, fra loro indipendenti. Allora

$$\mathbb{E}\{X\} = \mathbb{E}\{Y_1\} + \mathbb{E}\{Y_2^2\} = \delta^2 + n$$

e

$$\text{Var}\{Y\} = \text{Var}\{Y_1\} + \text{Var}\{Y_2^2\} = 2(n-1) + \text{Var}\{Y_2^2\}.$$

Siccome $Y_2 = \delta + (Y_2 - \delta)$, ricordando momento terzo e quarto di una variabile gaussiana

$$\text{Var}\{Y_2\} = \text{Var}\{(Y_2 - \delta)^2\} + \text{Var}\{2\delta(Y_2 - \delta)\} + 2\text{cov}\{(Y_2 - \delta)^2, 2\delta(Y_2 - \delta)\} = 2 + 4\delta^2,$$

quindi

$$\text{Var}\{X\} = 2n + 4\delta^2.$$

□

Supponiamo infine che $Y_1 \sim \chi^2(n_1, \delta_1)$ e $Y_2 \sim \chi^2(n_2, \delta_2)$ siano indipendenti, allora dalla definizione 5.3.1 segue immediatamente che:

$$Y_1 + Y_2 \sim \chi^2(n, \delta)$$

dove $n = n_1 + n_2$ e $\delta^2 = \delta_1^2 + \delta_2^2$.

La prossima proposizione determina la distribuzione di una forma quadratica di vettori gaussiani: essa sarà molto utile sia in questo che nei prossimi capitoli.

Proposizione 5.3.3. *Supponiamo che $Y^t = (Y_1, \dots, Y_n)$ sia un vettore gaussiano, con vettore delle medie $\mu = (\mu_1, \dots, \mu_n)$ e matrice di covarianza Q non singolare. Sia A una matrice di dimensione $n \times n$ simmetrica; allora:*

$$Y^t A Y \sim \sum_{i=1}^f \lambda_i X_i,$$

dove i λ_i sono gli f autovalori di QA , q_i è la molteplicità di λ_i e le variabili aleatorie X_i hanno legge $\chi^2(q_i, \delta_i)$, con $(\delta_1, \dots, \delta_f) \in \mathbb{R}^f$ funzione lineare di μ .

Dimostrazione. La matrice di covarianza Q è definita positiva, quindi esiste una matrice non singolare P tale che

$$PQP^t = I,$$

dove I indica la matrice identità di dimensione n . Consideriamo la matrice

$$(P^t)^{-1}AP^{-1},$$

essa è definita positiva, dunque esiste una matrice ortogonale T tale che

$$T(P^t)^{-1}AP^{-1}T^t = D,$$

dove D è una matrice diagonale con $D_{i,i} = \lambda_i$. Sia $Z = (Z_1, \dots, Z_n)$ il vettore aleatorio definito da

$$Z = (TP)Y$$

esso è un vettore gaussiano, perché Y lo è, con vettore delle medie e matrice di covarianza:

$$\mathbb{E}\{Z\} = (TP)\mathbb{E}\{Y\} \quad \text{Var}\{Z\} = (TP)Q(TP)^t = I,$$

perché T è ortogonale. Inoltre, visto che $P^{-1}T^tTP = I$

$$Y^tAY = Y^tP^tT^tT(P^t)^{-1}AP^{-1}T^tTPY = Z^tDZ = \sum_{i=1}^n \lambda_i Z_i^2.$$

I coefficienti λ_i sono gli autovalori della matrice

$$(P^t)^{-1}AP^t;$$

dal momento che gli autovalori sono invarianti per coniugio e $P = (P^t)^{-1}Q^{-1}$, i λ_i sono anche gli autovalori della matrice

$$P^{-1}PQAQP^t(P^t)^{-1}Q^{-1} = QA.$$

Quindi se raccogliamo gli addendi con uguale coefficiente λ_i

$$\sum_{i=1}^n \lambda_i Z_i^2 = \sum_{j=1}^f \lambda_j X_j,$$

dove f è il numero di autovalori distinti di QA , le variabili aleatorie X_i sono indipendenti fra loro con

$$X_i \sim \chi^2(q_i, \delta_i);$$

δ_i è la molteplicità dell'autovalore λ_i e il vettore δ_i è ottenuto tramite una trasformazione lineare da $\mathbb{E}\{Y\}$, dunque anche da μ . \square

Corollario 5.3.1. *Supponiamo che $Y = (Y_1, \dots, Y_n)^t$ sia un vettore gaussiano centrato, con matrice di covarianza Q non singolare, e sia A una matrice quadrata e simmetrica di dimensione n . Allora*

$$\mathbb{E}\{Y^t AY\} = \text{Tr}QA.$$

Dimostrazione. Segue direttamente dalla proposizione precedente. \square

Ci proponiamo di studiare il caso in cui la matrice di covarianza Q del vettore gaussiano $Y = (Y_1, \dots, Y_n)^t$ è singolare. Anche in questo caso esiste una matrice non singolare P tale che

$$PQP^t = \bar{I},$$

dove \bar{I} è così definita:

$$\bar{I} = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$$

e m è il rango di Q . Esiste poi una matrice ortogonale T tale che

$$T(P^t)^{-1}AP^tT^t = D,$$

dove D è la matrice diagonale con $D_{i,i} = \lambda_i$. Il vettore aleatorio

$$Z = (Z_1, \dots, Z_n),$$

definito da

$$Z = TPY$$

è un vettore gaussiano con vettore delle medie

$$\mathbb{E}Z = TP\mathbb{E}\{Y\}$$

e matrice di covarianza \bar{I} . Dunque $n - m$ componenti di Z sono quasi certamente costanti. Infine

$$Y^t AY = \sum_{i=1}^n \lambda_i Z_i^2 = k + \sum_{i=1}^f \lambda_i X_i,$$

dove k è una costante, che dipende dalla distribuzione di Y ; le variabili aleatorie X_i sono indipendenti ed hanno distribuzione $\chi^2(q_i, \delta_i)$ dove i λ_i sono gli autovalori della matrice

$$(P^t)^{-1}AP^t$$

e q_i è la molteplicità dell'autovalore λ_i . I δ_i , come prima, sono funzioni lineari di $\mathbb{E}\{Y\}$. In questo caso, tuttavia, visto che la matrice Q non è invertibile, non possiamo concludere che i λ_i siano gli autovalori di QA .

5.3.2 Stimatori per $\Gamma^*(n)$

D'ora in avanti, riprendiamo ad usare la terminologia introdotta nel paragrafo 5.1. Come fatto per la media, il primo problema che affrontiamo è quello dell'inferenza statistica su $\Gamma^*(n)$, a partire dalle misure relative ad un singolo individuo. Ricordiamo che la funzione Γ^* non è la funzione di covarianza, ma la funzione

$$\Gamma^*(t) = \Gamma(t) + \mu^2,$$

dove Γ è la funzione di covarianza del processo.

Dato $n \in \mathbb{N}$ con $m_i > n$, consideriamo le variabili aleatorie

$$D_i(n) = \frac{1}{m_i - n} \sum_{k=0}^{m_i - n - 1} Y_i(k)Y_i(k + m_i) :$$

esse sono stimatori corretti di $\Gamma^*(n)$. In questo paragrafo studiamo il caso particolare in cui $n = 0$; dimostrazioni simili potrebbero essere fatte per ogni n (complicando un po' i calcoli), purché si abbia $m_i > n$ per ogni i .

Per semplificare la notazione, d'ora in avanti al posto di $D_i(0)$ scriveremo semplicemente D_i . Nel lemma che segue studiamo la distribuzione di D_i ipotizzando che la funzione di covarianza Γ sia definita positiva.

Lemma 5.3.1. *Sia $\{Y(k)\}_{k \in \mathbb{N}}$ un processo stazionario gaussiano, con media μ e funzione di covarianza Γ definita positiva. Sia $m \in \mathbb{N}$; allora la variabile aleatoria:*

$$D = \frac{1}{m} \sum_{k=0}^m Y^2(k),$$

è uno stimatore corretto di $\Gamma(0) + \mu^2$ e

$$\text{Var}\{D\} \leq 4\Gamma(0)^2 + 3\Gamma(0)\mu^2.$$

Dimostrazione. La correttezza dello stimatore è ovvia: dimostriamo solo la seconda affermazione. La variabile aleatoria D è forma quadratica del vettore gaussiano $(Y(0), \dots, Y(m))$ con matrice associata la matrice identità I , di dimensione $m + 1$. Indichiamo con $Q = \{q_{i,j}\}_{i,j=0,\dots,m}$ la matrice

$$q_{i,j} = \Gamma(j - i).$$

Per l'ipotesi fatta su Γ , la matrice Q è definita positiva. Dalla proposizione 5.3.3 segue che

$$D \sim \frac{1}{m} \sum_{i=1}^f \lambda_i X_i$$

dove i λ_i sono gli autovalori della matrice Q e, se q_i è la molteplicità di λ_i e f è il numero di autovalori distinti di Q , le variabili X_i sono indipendenti con legge

$$X_i \sim \chi^2(q_i, \delta_i),$$

con $(\delta_1, \dots, \delta_f)$ funzione lineare di μ . Indichiamo con λ il raggio spettrale della matrice Q . Poiché gli autovalori di Q sono tutti positivi e la traccia di una matrice è invariante per coniugio, abbiamo:

$$\lambda \leq \sum_{i=1}^f q_i \lambda_i = \sum_{i=1}^m q_{i,i} = m\Gamma(0).$$

Siccome lo stimatore D è corretto, ricordando il valore atteso di una variabile χ^2 decentrata si ha:

$$\begin{aligned} \Gamma(0) + \mu^2 &= \mathbb{E}\{D\} = \frac{1}{m} \sum_{i=1}^f \lambda_i \mathbb{E}\{X_i\} = \\ &= \frac{1}{m} \sum_{i=1}^f \lambda_i q_i + \frac{1}{m} \sum_{i=1}^f \lambda_i \delta_i^2. \end{aligned}$$

Da ciò otteniamo che

$$\sum_{i=1}^f \lambda_i q_i + \sum_{i=1}^f \lambda_i \delta_i^2 = m\Gamma(0) + m\mu^2 \text{ e } \sum_{i=1}^f \lambda_i \delta_i^2 \leq m(\Gamma(0) + \mu^2).$$

Abbiamo gli strumenti necessari per dare una maggiorazione della varianza del nostro stimatore:

$$\begin{aligned} m^2 \text{Var}\{D\} &= \text{Var}\left\{\sum_{i=1}^f \lambda_i X_i\right\} = \sum_{i=1}^f \lambda_i^2 \text{Var}\{X_i\} = \\ &= \sum_{i=1}^f \lambda_i^2 (2q_i + 4\delta_i^2) \leq 2\lambda \left(\sum_{i=1}^f \lambda_i q_i + \sum_{i=1}^f \lambda_i \delta_i^2\right) + 2\lambda \left(\sum_{i=1}^f \lambda_i \delta_i^2\right). \end{aligned}$$

Per quanto detto sopra

$$2\lambda \left(\sum_{i=1}^f \lambda_i q_i + \sum_{i=1}^f \lambda_i \delta_i^2\right) + 2\lambda \left(\sum_{i=1}^f \lambda_i \delta_i^2\right) \leq 4m^2 \Gamma(0)^2 + 3m^2 \Gamma(0) \mu^2$$

e dunque

$$\text{Var}\{\hat{\Gamma}_m(0)\} \leq 4\Gamma(0)^2 + 3\Gamma(0)\mu^2.$$

□

Per ogni soggetto abbiamo dunque definito uno stimatore D_i . Poiché abbiamo dimostrato che questo è uno stimatore corretto e ipotizziamo che i processi relativi a diversi pazienti siano fra loro indipendenti, costruiamo uno stimatore globale prendendo la media campionaria dei D_i ; definiamo quindi la successione di stimatori

$$E_n = \frac{1}{n} \sum_{i=1}^n D_i.$$

Condizione base per una successione di stimatori di un parametro è che essa sia, in qualche senso, consistente. In questo caso, la successione $\{E_n\}_{n \in \mathbb{N}}$ è una successione fortemente consistente di stimatori di $\Gamma^*(0)$: questo deriva immediatamente dal lemma 5.3.1. Infatti, dal momento che

$$\text{Var}\{D_i\} \leq 4\Gamma(0)^2 + 3\Gamma(0)\mu^2,$$

si ha

$$\sum_{i=1}^{\infty} \frac{\text{Var}\{D_i\}}{i^2} < \infty.$$

Quindi sono soddisfatte le ipotesi del teorema 5.2.1 che garantisce che la successione di variabili aleatorie $\{E_n\}_{n \in \mathbb{N}}$ converga quasi certamente a

$$\mathbb{E}\{E_n\} = \Gamma(0) + \mu^2.$$

5.3.3 Inferenza statistica sulla funzione di covarianza nel caso markoviano

In questa sezione supponiamo che i processi

$$\{Y_i(n)\}_{t \in \mathbb{N}}$$

siano di Markov: per quanto in 3.3, la funzione di covarianza Γ è dunque nella forma

$$\Gamma(k) = \Gamma(0)\theta^{|k|} \quad (5.4)$$

per $k \in \mathbb{Z}$, dove

$$\theta = \frac{\Gamma(1)}{\Gamma(0)}.$$

In questo caso, dunque, data la parametrizzazione (5.4) è possibile un'inferenza sull'intera funzione Γ , anziché solo sui valori che essa assume in particolari punti.

Il metodo statistico più comune per determinare degli stimatori è quello della massima verosimiglianza: qui spieghiamo perché non è stato adottato in questo caso. Poniamoci nel caso più semplice in cui tale stima sia possibile: supponiamo di avere solo 2 osservazioni per ogni paziente e, fissato un $i \in \mathbb{N}$, consideriamo il modello statistico indotto dal campione

$$(Y_i(0), Y_i(1)).$$

Sotto l'ipotesi che valga la parametrizzazione (5.4), il campione è un vettore gaussiano, con matrice Q di covarianza definita da

$$Q = \Gamma(0) \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix},$$

la cui inversa è data da

$$Q^{-1} = \Gamma(0)^{-1}(1 - \theta^2)^{-1} \begin{pmatrix} 1 & -\theta \\ -\theta & 1 \end{pmatrix}.$$

Quindi la verosimiglianza del modello statistico in considerazione rispetto alla misura di Lebesgue su \mathbb{R}^2 , data da

$$L(y_1, y_2, \mu, \theta, \Gamma(0)) = \frac{1}{\sqrt{2\pi \det Q}} \exp\{(y_1 - \mu, y_2 - \mu)Q^{-1}(y_1 - \mu, y_2 - \mu)^t\},$$

può essere sviluppata come

$$L(y_1, y_2, \mu, \theta, \Gamma(0)) = \frac{1}{2\pi} \frac{1}{\Gamma(0)(1 - \theta^2)} \exp\{f(y_1, y_2, \mu, \theta, \Gamma(0))\},$$

dove

$$f(y_1, y_2, \mu, \theta, \Gamma(0)) = \frac{1}{\Gamma(0)(1 - \theta^2)} [(y_1 - \mu)^2 + (y_2 - \mu)^2 - 2\theta(y_1 - \mu)(y_2 - \mu)].$$

Fissate le osservazioni $y_1, y_2 \in \mathbb{R}$, la funzione di verosimiglianza g è di classe C^2 sull'aperto

$$(\mu, \Gamma(0), \theta) \in \mathbb{R} \times \mathbb{R}^+ \times (0, 1);$$

per trovare un eventuale punto di massimo della verosimiglianza, dunque, potremmo risolvere le equazioni di *score*, ottenute imponendo che si annulli il gradiente di L . Nel nostro caso, tuttavia, il sistema

$$\nabla L = 0$$

non sembra condurre a soluzioni di forma analitica chiusa: dunque, neanche in questo caso, che è il più semplice che si possa considerare (perché è basato solo su due osservazioni), gli stimatori di massima verosimiglianza sono di facile derivazione. In molti testi statistici, fra cui [31], in casi come questi si procede ad una stima numerica del punto di massimo per la verosimiglianza. Questo procedimento, spesso usato, deve essere integrato con alcune considerazioni sul modello statistico preso in considerazione, perché gli stimatori di massima verosimiglianza sono asintoticamente normali e costituiscono una successione fortemente consistente per i parametri solo in determinate famiglie di modelli, come quelli esponenziali. Quindi abbandoniamo questo metodo di stima e cerchiamo in altro modo degli stimatori fortemente consistenti.

Poiché abbiamo già dato, nei paragrafi precedenti, degli stimatori di μ e, indirettamente, di $\Gamma(0)$, ci occupiamo di determinare una successione di stimatori fortemente consistente di θ . Abbiamo una successione di stimatori fortemente consistente per $\Gamma(0)$, una successione di stimatori fortemente consistente per $\Gamma(1)$, dal momento che

$$\theta = \frac{\Gamma(1)}{\Gamma(0)},$$

combinando queste due successioni avremo (ove essa risulti ben definita) una successione di stimatori fortemente consistente di θ . La successione proposta è dunque

$$\frac{D_n(1) - C_n^2}{D_n(0) - C_n^2}$$

dove le successioni $D_n(h)$ e C_n sono state definite in questo capitolo, nelle sezioni 2 e 3.

5.4 Risultati ottenuti

Riportiamo i risultati ottenuti con i dati che abbiamo a disposizione. Si possono leggere nella tabella 5.1: sotto ai valori stimati è riportata la taglia del campione che è stato utilizzato per calcolare tali stime.

Tabella 5.1: Stime ottenute.

$\hat{\mu}$	$\hat{\Gamma}(0)$	$\hat{\Gamma}(1)$	$\hat{\Gamma}(2)$	$\hat{\Gamma}(3)$	$\hat{\Gamma}(4)$
12.597	2.4625	1.59967	1.9246	1.7608	1.9525
2548	1985	1899	1543	987	677

I valori sono stati stimati su campioni di taglia differente perché i pazienti non si sono presentati sempre a tutte le visite. Da questi dati ricaviamo anche lo stimatore dato per θ nel caso in cui il processo sia di Markov. Se indichiamo con $\Gamma(0)$ e $\Gamma(1)$ le stime appena descritte, esso è dato da:

$$\frac{\Gamma(1)}{\Gamma(0)} = 0.64961.$$

Possiamo notare, fra l'altro, che se il processo fosse di Markov il rapporto

$$\frac{\Gamma(i+1)}{\Gamma(i)} = \theta$$

dovrebbe essere costante, cosa che non sembra essere evidenziata dai nostri stimatori. Questo è solo un suggerimento, ma potrebbe essere testata rigorosamente con appositi test di markovianità, che noi non considereremo perché ci sono sufficienti i valori di $\Gamma(n)$ stimati in precedenza.

Capitolo 6

Un approccio alternativo alla stima dei parametri nei sani.

Consideriamo il processo trasformato $\{f_\lambda(X(t))\}_{t \in \mathbb{R}^+}$, di cui abbiamo discusso precedentemente. Anche in questo capitolo ci proponiamo di dare degli stimatori della media e della funzione di covarianza di questo processo, basati su di un approccio alternativo.

Riconduciamo i dati di cui disponiamo alle osservazioni delle prime 45 variabili di un processo gaussiano stazionario $\{Y(n)\}_{n \in \mathbb{N}}$, con media e funzione di covarianza uguali (a meno di un fattore moltiplicativo) a quelle del precedente.

La stima della media utilizzerà, sostanzialmente, gli stessi stimatori introdotti nel capitolo precedente. Per la stima della funzione di covarianza, invece, usiamo particolari stimatori, detti *correlogrammi*. A costo di imporre particolari condizioni sulla distribuzione spettrale o sulla struttura del processo, riusciremo a provare che questi stimatori hanno buone proprietà asintotiche.

Si tratta comunque di risultati più deboli rispetto a quelli ottenuti nel capitolo precedente: le successioni di stimatori sono consistenti in probabilità, o sono solo asintoticamente corrette.

Non testiamo in alcun modo le ipotesi sulla distribuzione spettrale o sulla struttura del processo; abbiamo a disposizione, per giunta, soltanto 45 osservazioni, mentre i risultati hanno valore asintotico: per queste ragioni, nel proseguire la nostra analisi, non utilizzeremo le stime date in questo capitolo, ma quelle che sono già state ricavate nel capitolo 5.

6.1 Notazione

Riprendiamo la notazione utilizzata nel capitolo precedente: il decadimento del logaritmo dei linfociti $cd4+$ per mm^3 di sangue nell' i -esimo soggetto sano è modellizzato da un processo gaussiano a tempi discreti

$$\{Y_i(n)\}_{n \in \mathbb{N}}.$$

I processi relativi a pazienti diversi hanno uguali proprietà statistiche e sono equidistribuiti. Nel capitolo precedente abbiamo analizzato i risultati asintotici degli stimatori ricavati per $i \rightarrow \infty$. In questo capitolo procediamo in maniera diversa: vogliamo dare dei risultati asintotici su n , ovvero sul numero di osservazioni di ciascun processo.

Consideriamo perciò i pazienti che si sono presentati ad ogni visita; sono 198. A meno di riordinare l'insieme degli indici, possiamo supporre che siano i pazienti indicizzati da i , con $i \leq 198$; consideriamo allora il processo gaussiano e stazionario $\{Y(n)\}_{n \in \mathbb{N}}$, definito da:

$$Y(n) = \frac{1}{198} \sum_{i=1}^{198} Y_i(n).$$

Esso ha media μ uguale alla media dei singoli processi $Y_i(n)$, mentre la sua funzione di covarianza Γ è definita da

$$\Gamma(h) = \frac{1}{198} \text{cov}\{Y_i(n), Y_i(n+h)\}.$$

In questo capitolo indicheremo sempre con $Y(n)$ questo processo, sottintendendo quindi che sia gaussiano e stazionario.

6.2 La stima della media

In questo paragrafo vogliamo determinare degli stimatori per la media μ del processo stazionario $\{Y(k)\}_{k \in \mathbb{N}}$ basati sulle osservazioni delle variabili aleatorie $Y(0), \dots, Y(n)$. Il problema è analogo a quello che abbiamo affrontato nel capitolo 5 per la stima di μ basandosi sui dati di un singolo soggetto. Anche in questo caso scegliamo la media campionaria

$$\bar{Y}_n = \frac{1}{n+1} \sum_{i=0}^n Y(i).$$

Nel paragrafo 5.2.2 è stato osservato che, nel modello statistico indotto dal campione

$$Y(0), \dots, Y(n),$$

esiste uno stimatore di μ preferibile a questo e sono state indicate le motivazioni che hanno portato a non adottarlo. Tali motivazioni rimangono valide anche in questa situazione.

Analizziamo le proprietà asintotiche della successione di stimatori $\{\bar{Y}_n\}_{n \in \mathbb{N}}$. La proprietà minimale che si richiede ad una successione di stimatori di un parametro è la consistenza di qualche genere.

La proposizione 6.2.1 fornisce delle condizioni necessarie e sufficienti affinché la successione delle medie campionarie sia una successione di stimatori di μ consistente in $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Anticipiamo due lemmi.

Lemma 6.2.1. *Consideriamo il processo $\{Y(n)\}_{n \in \mathbb{N}}$. Sia ν la distribuzione spettrale del processo; se essa è continua in $\{0\}$ la successione \bar{Y}_n converge a μ in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.*

Dimostrazione. A meno di considerare il processo $Y(n) - \mu$, possiamo supporre che $Y(n)$ sia un processo centrato. Per il teorema di rappresentazione spettrale di Kolmogorov esiste un processo stocastico $\{Z(t)\}_{t \in (-\pi, \pi]}$ ad incrementi ortogonali tale che:

$$Y(n) = \int_{(-\pi, \pi]} e^{ik\lambda} Z(d\lambda).$$

Definiamo la successione di funzioni $\{f_n\}_{n \in \mathbb{N}}$ come:

$$f_n(x) = \begin{cases} 1 & \text{se } x = 0; \\ \frac{1 - e^{i(n+1)x}}{(n+1)(1 - e^{ix})} & \text{se } x \neq 0. \end{cases}$$

Puntualmente, per $n \rightarrow \infty$, la successione f_n converge all'indicatrice di 0, I_0 che è uguale, a meno di insiemi di probabilità nulla, alla funzione 0 se e solo se $\nu\{0\} = 0$.

Poiché $\|f_n\| \leq 1$ per ogni n , il teorema della convergenza dominata implica che questa convergenza avvenga anche in media quadratica. Poiché

$$\bar{Y}_n = \frac{\sum_{i=0}^n Y(i)}{n+1} = \int_{(-\pi, \pi]} \frac{1}{n+1} \sum_{k=0}^n e^{ikx} Z(d\lambda) = \int_{(-\pi, \pi]} f_n(\lambda) Z(d\lambda),$$

passando al limite su $n \rightarrow \infty$ si ha:

$$\lim_n \bar{Y}_n = \int_{(-\pi, \pi]} I_0(\lambda) Z(d\lambda).$$

Dunque, se $\nu\{0\} = 0$, allora

$$\lim_n \bar{Y}_n = 0.$$

□

Lemma 6.2.2. *Consideriamo il processo $\{Y(n)\}_{n \in \mathbb{N}}$; sia ν la misura spettrale di $\{Y(n)\}_{n \in \mathbb{N}}$ e Γ la sua funzione di covarianza. Condizione necessaria e sufficiente affinché ν sia continua in $\{0\}$ è che*

$$\lim_n \frac{1}{n+1} \sum_{k=0}^n \Gamma(k) = 0.$$

Dimostrazione. Per il lemma di Herglotz si ha

$$\Gamma(k) = \int_{(-\pi, \pi]} e^{ikx} d\nu(x).$$

Definiamo, come nella dimostrazione precedente, le funzioni f_n come

$$f_n(x) = \begin{cases} 1 & \text{se } x = 0; \\ \frac{1 - e^{i(n+1)x}}{(n+1)(1 - e^{ix})} & \text{se } x \neq 0. \end{cases}$$

Come prima, esse convergono puntualmente a I_0 , indicatrice di $\{0\}$; inoltre in $((-\pi, \pi], \mathcal{B}(0, 1], \nu)$ la funzione I_0 è uguale, a meno di insiemi di misura nulla, alla costante 0 se e solo se la misura spettrale ν è continua in $\{0\}$. Dal momento che

$$\frac{1}{n+1} \sum_{k=0}^n \Gamma(k) = \int_{(-\pi, \pi]} \sum_{k=0}^n e^{ikx} d\nu = \int_{(-\pi, \pi]} f_n(x) d\nu.$$

e che $\|f_n\| \leq 1$, per il teorema della convergenza dominata si ha

$$\lim_n \frac{1}{n+1} \sum_{k=0}^n \Gamma(k) = \int_{(-\pi, \pi]} I_0(x) d\nu(x). \quad (6.1)$$

e il membro destro di (6.1) è nullo se e solo se ν è continua in $\{0\}$. □

Proposizione 6.2.1. Consideriamo il processo $\{Y(n)\}_{n \in \mathbb{N}}$. La condizione

$$\lim_n \frac{1}{n+1} \sum_{k=0}^n \Gamma(k) = 0$$

è necessaria e sufficiente affinché la successione $\{\bar{Y}_n\}_{n \in \mathbb{N}}$ converga in media quadratica a μ .

Dimostrazione. La sufficienza della condizione deriva immediatamente dai lemmi 6.2.1 e 6.2.2. Dimostriamo che è anche una condizione necessaria. Per la disuguaglianza triangolare

$$\begin{aligned} \left| \frac{1}{n+1} \sum_{k=0}^n \Gamma(k) \right| &= \left| \mathbb{E} \left\{ \frac{1}{n+1} \sum_{k=0}^n (Y(k) - \mu)(Y(0) - \mu) \right\} \right| \leq \\ &< \mathbb{E} \{ |Y(0) - \mu|^2 \} \mathbb{E} \left\{ \left| \frac{1}{n+1} \sum_{k=0}^n Y(k) - \mu \right|^2 \right\}. \end{aligned}$$

Quindi

$$\frac{1}{n+1} \sum_{k=0}^n Y(k) \xrightarrow{L^2} \mu$$

implica che

$$\lim_n \frac{1}{n+1} \sum_{k=0}^n \Gamma(k) = 0.$$

□

6.3 La stima della funzione di covarianza: i correlogrammi

In questa sezione determineremo degli stimatori per alcuni valori della funzione di covarianza. Consideriamo, per $m \in \mathbb{N}$, la successione di variabili aleatorie

$$\hat{\Gamma}_n(m) = \frac{1}{n+1} \sum_{i=0}^{n-m} (Y(i) - \bar{Y}_n)(Y(i+m) - \bar{Y}_n),$$

definita per $n > m$, dove \bar{Y}_n è la media campionaria

$$\bar{Y}_n = \frac{1}{n+1} \sum_{i=0}^n Y(i). \quad (6.2)$$

Le variabili $\hat{\Gamma}_n(m)$ si dicono solitamente *correlogrammi*. Essi non sono degli stimatori corretti di $\Gamma(m)$, tuttavia, ponendo opportune ipotesi sul processo $\{Y(n)\}_{n \in \mathbb{N}}$, sono dotati di buone proprietà asintotiche.

I correlogrammi del processo $\{Y(n)\}_{n \in \mathbb{N}}$ coincidono con i correlogrammi del processo centrato $\{Y(n) - \mu\}_{n \in \mathbb{N}}$. Per questo possiamo supporre, senza perdere di generalità, che il processo $\{Y(n)\}_{n \in \mathbb{N}}$ sia centrato.

Siccome poi operare sui correlogrammi rende la notazione molto complessa, consideriamo le variabili aleatorie

$$Z_n(m) = \frac{1}{n+1} \sum_{i=0}^n Y(i)Y(i+m). \quad (6.3)$$

La differenza $Z_n(m) - \hat{\Gamma}_n(m)$ può essere sviluppata come

$$\bar{Y}_n \left[\frac{1}{n+1} \sum_{i=0}^{n-m} Y(i+m) + \frac{1}{n+1} \sum_{i=0}^{n-m} Y(i) - \left(1 - \frac{m}{n+1}\right) \bar{Y}_n \right] + \frac{1}{n+1} \sum_{i=n-m+1}^n Y(i)Y(i+m) \quad (6.4)$$

Il lemma che segue ci autorizzerà a dimostrare i risultati asintotici dei correlogrammi usando la successione di variabili $\{Z_n(m)\}_{n \in \mathbb{N}}$. Enunciamo prima il teorema di Slutsky per somma e prodotto, che servirà nella dimostrazione del lemma (per una dimostrazione, vedi [26]).

Teorema 6.3.1. *Supponiamo che $\{Y_n\}_{n \in \mathbb{N}}$ e $\{X_n\}_{n \in \mathbb{N}}$ siano due successioni di variabili aleatorie definite su $(\Omega, \mathcal{F}, \mathbb{P})$, tali che*

$$Y_n \xrightarrow{\mathcal{L}} Y \quad X_n \xrightarrow{\mathbb{P}} 0,$$

allora

$$Y_n \cdot X_n \xrightarrow{\mathbb{P}} 0 \quad e \quad X_n + Y_n \xrightarrow{\mathcal{L}} Y.$$

Lemma 6.3.1. *Supponiamo che il processo gaussiano $\{Y(n)\}_{n \in \mathbb{N}}$ sia centrato. Siano $\hat{\Gamma}_n(m)$ e $Z_n(m)$ le variabili aleatorie definite rispettivamente da (6.2) e da (6.3), allora*

a) Se la funzione di covarianza è assolutamente sommabile, cioè se

$$\sum_{k=-\infty}^{\infty} |\Gamma(k)| < \infty,$$

allora

$$\sqrt{n+1}(Z_n(m) - \hat{\Gamma}_n(m)) \xrightarrow{\mathbb{P}} 0.$$

b) Indichiamo con ν la distribuzione spettrale di $\{Y(n)\}_{n \in \mathbb{N}}$. Se ν è continua in $\{0\}$, allora per $n \rightarrow \infty$

$$Z_n(m) - \hat{\Gamma}_n(m) \xrightarrow{\mathbb{P}} 0.$$

Dimostrazione. a). Per l'uguaglianza (6.4), la differenza

$$\sqrt{n+1}(Z_n(m) - \hat{\Gamma}_n(m))$$

può essere scomposta nella somma delle due successioni

$$\sqrt{n+1} \bar{Y}_n \left[\frac{1}{n+1} \sum_{i=0}^{n-m} Y(i+m) + \frac{1}{n+1} \sum_{i=0}^{n-m} Y(i) - \left(1 - \frac{m}{n+1}\right) \bar{Y}_n \right] \quad (6.5)$$

e

$$\frac{1}{\sqrt{n+1}} \sum_{i=n-m+1}^n Y(i)Y(i+m). \quad (6.6)$$

Ci basta dimostrare che entrambe le successioni convergono alla costante 0 in probabilità.

Dimostriamo la convergenza di (6.5). Poiché il processo $\{Y(n)\}_{n \in \mathbb{N}}$ è gaussiano

$$\sqrt{n+1} \bar{Y}_n \sim N\left(0, \sum_{|h| < n+1} \left(1 - \frac{|h|}{n+1}\right) \Gamma(h)\right).$$

Per ipotesi la funzione di covarianza è assolutamente sommabile, dunque si ha

$$\lim_n \text{Var}\{\sqrt{n+1} \bar{Y}_n\} = \sum_{h=-\infty}^{\infty} \Gamma(h).$$

Dal momento che la successione

$$\sqrt{n+1} \bar{Y}_n$$

è una successione di variabili gaussiane con medie e varianze convergenti a valori finiti, il teorema di Lévy garantisce che vale la convergenza in legge

$$\sqrt{n+1} \bar{Y}_n \xrightarrow{\mathcal{L}} Y$$

con $Y \sim N(0, \sum_{h=-\infty}^{\infty} \Gamma(h))$.

Se la funzione di covarianza è assolutamente sommabile sono verificate le ipotesi della proposizione 6.2.1, perché, per l'assoluta sommabilità della funzione di covarianza, vale

$$\lim_{k \rightarrow \infty} |\Gamma(k)| = 0,$$

ma allora per il teorema di Césaro

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n |\Gamma(k)| = 0,$$

da cui si deduce

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \Gamma(k) = 0.$$

Per la proposizione 6.2.1, quindi \mathbb{P} -quasi certamente,

$$\frac{1}{n+1} \sum_{i=0}^{n-m} Y(i+m) + \frac{1}{n+1} \sum_{i=0}^{n-m} Y(i) - \left(1 - \frac{m}{n+1}\right) \bar{Y}_n \rightarrow 0.$$

La successione di variabili aleatorie (6.5) è prodotto di due successioni di variabili aleatorie, una convergente in legge e una convergente in probabilità a zero. Per il teorema 6.3.1 per il prodotto abbiamo quindi:

$$\sqrt{n+1} \bar{Y}_n \left[\frac{1}{n+1} \sum_{i=0}^{n-m} Y(i+m) + \frac{1}{n+1} \sum_{i=0}^{n-m} Y(i) - \left(1 - \frac{m}{n+1}\right) \bar{Y}_n \right] \xrightarrow{\mathbb{P}} 0$$

Dimostriamo la convergenza della variabile aleatoria (6.6). Dal momento che la funzione di covarianza verifica

$$|\Gamma(h)| \leq \Gamma(0),$$

per la disuguaglianza triangolare si ha:

$$\frac{1}{\sqrt{n+1}} \mathbb{E}\left\{ \left| \sum_{i=n-m+1}^n Y(i)Y(i+m) \right| \right\} \leq \frac{m\Gamma(0)}{\sqrt{n+1}},$$

dunque la variabile aleatoria

$$\frac{1}{\sqrt{n+1}} \sum_{i=n-m+1}^n Y(i)Y(i+m)$$

converge in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ a 0, quindi anche in probabilità.

b). Consideriamo ancora la decomposizione della successione di variabili aleatorie $\hat{\Gamma}_n(m) - Z_n(m)$ nella somma delle due successioni

$$\bar{Y}_n \left[\frac{1}{n+1} \sum_{i=0}^{n-m} Y(i+m) + \frac{1}{n+1} \sum_{i=0}^{n-m} Y(i) - \left(1 - \frac{m}{n+1}\right) \bar{Y}_n \right] \quad (6.7)$$

e

$$\frac{1}{n+1} \sum_{i=n-m+1}^n Y(i)Y(i+m). \quad (6.8)$$

Si tratta di provare che entrambe convergono a zero in probabilità. Se la misura spettrale ν è continua, la proposizione 6.2.1 assicura che

$$\frac{1}{n+1} \sum_{i=0}^{n-m} Y(i+m) + \frac{1}{n+1} \sum_{i=0}^{n-m} Y(i) - \left(1 - \frac{m}{n}\right) \bar{Y}_n \xrightarrow{\mathbb{P}} 0$$

e che

$$\bar{Y}_n \xrightarrow{\mathbb{P}} 0.$$

Dunque la successione (6.7) converge a zero in probabilità. Consideriamo il valore atteso della seconda successione (6.8): in modo del tutto analogo a quanto fatto nel punto **a)** si dimostra che

$$\frac{1}{n+1} \mathbb{E} \left\{ \left| \sum_{i=n-m+1}^n Y(i)Y(i+m) \right| \right\} \leq \frac{m\Gamma(0)}{n+1};$$

quindi la successione di stimatori (6.8) converge a 0 in $L^1(\Omega, \mathcal{F}, \mathbb{P})$, dunque anche in probabilità. \square

Seguiamo due approcci diversi per derivare risultati asintotici sui correlogrammi.

Nel paragrafo 6.3.1, imposte delle condizioni sulla distribuzione spettrale del processo $\{Y(n)\}_{n \in \mathbb{N}}$, dimostreremo che i correlogrammi costituiscono una successione di stimatori di $\Gamma(m)$ debolmente consistente.

Nel paragrafo 6.3.2, invece, supporremo che il processo $\{Y(n)\}_{n \in \mathbb{N}}$ sia ottenuto mediante l'applicazione di un filtro lineare invariante nel tempo da un processo a variabili indipendenti; con questa ipotesi dimostreremo dei risultati di normalità asintotica per i correlogrammi.

6.3.1 Condizioni sulla distribuzione spettrale del processo.

Nella proposizione che segue enunciamo delle condizioni sulla distribuzione spettrale del processo $\{Y(n)\}_{n \in \mathbb{N}}$ sufficienti perché i correlogrammi siano una successione di stimatori di $\Gamma(m)$ debolmente consistente.

Proposizione 6.3.1. *Supponiamo che il processo gaussiano $\{Y(n)\}_{n \in \mathbb{N}}$ sia centrato. Indichiamo con Γ la funzione di covarianza di $\{Y(n)\}_{n \in \mathbb{N}}$ e con ν la misura spettrale di Γ . Se la misura ν è continua, per ogni $m \in \mathbb{N}$ vale*

$$\hat{\Gamma}_n(m) \xrightarrow{\mathbb{P}} \Gamma(m),$$

dove la successione di variabili aleatorie $\{\hat{\Gamma}_n(m)\}_{n \in \mathbb{N}}$ è definita come in 6.2.

Dimostrazione. Per il lemma di Herglotz si ha

$$\Gamma(k) = \int_{(-\pi, \pi]} e^{ikx} d\nu,$$

dunque per il teorema di Fubini Tonelli

$$\Gamma^2(k) = \int_{(-\pi, \pi]} \int_{(-\pi, \pi]} e^{ik(x-y)} d\nu(x) d\nu(y).$$

Definiamo le funzioni

$$\phi_n(x, y) = \begin{cases} 1 & \text{se } x = y; \\ \frac{1 - e^{i(n+1)(x-y)}}{(n+1)(1 - e^{i(x-y)})} & \text{se } x \neq y. \end{cases}$$

La successione di funzioni ϕ_n è dominata, perché $\|\phi_n\| \leq 1$ e converge puntualmente a I_A , dove l'insieme A è definito da

$$A = \{(x, y) \in \mathbb{R}^2 \mid x = y\}.$$

Dunque, poiché $\frac{1}{n+1} \sum_{k=0}^n e^{ik(x-y)} = \phi_n(x, y)$,

$$\begin{aligned} \lim_n \frac{1}{n+1} \sum_{k=0}^n \Gamma^2(k) &= \lim_n \int_{(-\pi, \pi]} \int_{(-\pi, \pi]} \phi_n(x, y) d\nu(x) d\nu(y) = \\ &= \int_{(-\pi, \pi]} \lim_n \phi_n(x, y) d\nu(x) d\nu(y) = \int_{(-\pi, \pi]} \nu(\{x\}) d\nu(x) = 0, \end{aligned} \quad (6.9)$$

dal momento che la misura è continua, duque $\nu(\{x\}) = 0$ per ogni $x \in (-\pi, \pi]$.

Fissato $m \in \mathbb{N}$, consideriamo il processo stocastico

$$\{Y(n)Y(n+m)\}_{n \in \mathbb{N}},$$

esso è stazionario in senso stretto, dal momento che $Y(n)$ è un processo stazionario in senso stretto. Dal momento che il processo è gaussiano $\mathbb{E}\{Y(n)^4\} < \infty$, quindi

$$\mathbb{E}\{Y(n+m)^2 Y(n)^2\} < \infty.$$

Per la proposizione 6.2.1, applicata al processo gaussiano e centrato $\{Y(n)Y(n+m) - \Gamma(m)\}_{n \in \mathbb{N}}$, la condizione

$$\lim_m \frac{1}{m+1} \sum_{i=0}^n \mathbb{E}\{(Y(n)Y(n+i) - \Gamma(m))(Y(n)Y(0) - \Gamma(m))\} \rightarrow 0 \quad (6.10)$$

è una condizione necessaria e sufficiente affinché

$$\lim_n Z_n(m) \rightarrow \Gamma(m),$$

dove le variabili $Z_n(m)$ sono definite da

$$Z_n(m) = \frac{1}{n+1} \sum_{i=0}^n Y(i)Y(i+m).$$

Poiché il processo è gaussiano:

$$\mathbb{E}\{(Y(n)Y(n+m) - \Gamma(m))(Y(n)Y(0) - \Gamma(m))\} = \Gamma^2(m) + \Gamma(m+n)\Gamma(m-n).$$

Dal momento che

$$|\Gamma(m+n)\Gamma(m-n)| \leq \Gamma^2(n+m) + \Gamma^2(n-m),$$

per (6.9), fissato n

$$\lim_{n \rightarrow \infty} \frac{1}{m+1} \sum_{m=0}^n (\Gamma^2(m) + \Gamma(m+n)\Gamma(m-n)) \rightarrow 0,$$

dunque, \mathbb{P} -quasi certamente,

$$Z_n(m) \rightarrow \Gamma(m)$$

e quindi tale convergenza avviene anche in probabilità. ν è continua, dunque è continua in 0: per la proposizione 6.3.1

$$\hat{\Gamma}_n(m) \xrightarrow{\mathbb{P}} \Gamma(m)$$

se e solo se

$$Z_n \xrightarrow{\mathbb{P}} \Gamma(m).$$

□

6.3.2 Ipotesi particolari sulla struttura del processo

In questo paragrafo, poste delle particolari ipotesi sul processo $\{Y(n)\}_{n \in \mathbb{N}}$, ricaveremo dei risultati di normalità asintotica per i correlogrammi. Specifichiamo da subito le ipotesi che rimarranno valide per tutto il paragrafo.

Ipotesi. Supporremo che esistano un processo $\{X(n)\}_{n \in \mathbb{Z}}$ centrato a variabili indipendenti e una successione reale $\{\phi_j\}_{j \in \mathbb{Z}}$, con

$$\mathbb{E}\{X(n)^2\} = \sigma^2, \quad \mathbb{E}\{X(n)^4\} = \theta\sigma^2 < \infty \text{ e } \sum_{j=-\infty}^{\infty} |\psi_j| < \infty,$$

tali che

$$Y(n) = \sum_{j=-\infty}^{\infty} \psi_j X(n-j). \quad (6.11)$$

Abbiamo già visto, nel paragrafo 3.2, che la serie (6.11) converge in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Anche in questo caso, preferiremmo studiare le proprietà asintotiche della successione $Z_n(m)$, definita in (6.3), anziché lavorare direttamente sui correlogrammi. Siamo autorizzati a farlo per il lemma seguente.

Lemma 6.3.2. *Siano definite le successioni di variabili aleatorie $\{\hat{\Gamma}_n(m)\}_{n \in \mathbb{N}}$ e $Z_n(m)$ rispettivamente come in (6.2) e in (6.3), se Z è una variabile aleatoria allora*

$$\sqrt{n}(Z_n(m) - \Gamma(m)) \xrightarrow{\mathcal{L}} Z$$

se e solo se

$$\sqrt{n}(\hat{\Gamma}_n(m) - \Gamma(m)) \xrightarrow{\mathcal{L}} Z.$$

Dimostrazione. Consideriamo la funzione di covarianza Γ del processo $\{Y(n)\}_{n \in \mathbb{N}}$. Per (6.11) si ha

$$\Gamma(h) = \sum_{j,k=-\infty}^{\infty} \psi_j \psi_k \mathbb{E}\{X(n-j)X(n+h-k)\}.$$

Poiché il processo $\{X(n)\}_{n \in \mathbb{Z}}$ è, per ipotesi, a variabili indipendenti, vale

$$\Gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}.$$

Visto che la successione ψ_j è assolutamente sommabile, lo è anche la funzione di covarianza Γ , cioè

$$\sum_{k=-\infty}^{\infty} |\Gamma(k)| < \infty,$$

quindi $\lim_{k \rightarrow \infty} |\Gamma(k)| = 0$ e, per il teorema di Césaro

$$\lim_n \frac{1}{n+1} \sum_{k=0}^n |\Gamma(k)| = 0,$$

da cui

$$\lim_n \frac{1}{n+1} \sum_{k=0}^n \Gamma(k) = 0.$$

Sono soddisfatte le ipotesi del punto **a)** della proposizione 6.3.1, per cui vale

$$\sqrt{n+1}(Z_n(m) - \hat{\Gamma}_n(m)) \xrightarrow{\mathbb{P}} 0.$$

Supponiamo che

$$\sqrt{n+1}(\hat{\Gamma}_n(m) - \Gamma(m)) \xrightarrow{\mathcal{L}} Z,$$

abbiamo che

$$\sqrt{n+1}(Z_n(m) - \Gamma(m)) = \sqrt{n+1}(\hat{\Gamma}_n(m) - \Gamma(m)) + \sqrt{n+1}(Z_n(m) - \hat{\Gamma}_n(m));$$

per il teorema 6.3.1, dunque

$$\sqrt{n+1}(Z_n(m) - \Gamma(m)) \xrightarrow{\mathcal{L}} Z.$$

L'implicazione opposta è del tutto analoga. \square

Cominciamo a dedurre la distribuzione asintotica delle variabili $Z_n(m)$.

Proposizione 6.3.2. *Consideriamo le variabili aleatorie*

$$Z_n(m) = \frac{1}{n+1} \sum_{i=0}^n Y(i)Y(i+m),$$

siano $m_1, m_2 \in \mathbb{N}$, allora vale

$$\begin{aligned} \lim_n (n+1) \text{cov}\{Z_n(m_1), Z_n(m_2)\} &= \\ &= (\theta - 3)\Gamma(m_1)\Gamma(m_2) + \sum_{j=-\infty}^{\infty} [\Gamma(j)\Gamma(j - m_1 + m_2) + \Gamma(j + m_1)\Gamma(j - m_2)]. \end{aligned}$$

Dimostrazione. Poiché il processo $\{X(n)\}_{n \in \mathbb{Z}}$ è a variabili indipendenti, si ha

$$\mathbb{E}\{X(s)X(t)X(u)X(v)\} = \begin{cases} \theta\sigma^4 & \text{se } s = t = u = v \\ \sigma^4 & \text{se gli indici sono uguali a coppie, ma non tutti uguali} \\ 0 & \text{nei casi restanti.} \end{cases} \quad (6.12)$$

La serie (6.11) converge in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, quindi per $h \in \mathbb{Z}$, anche la serie

$$\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \psi_i \psi_{j+m_1} \psi_{k+h+m_1} \psi_{l+h+m_1+m_2} X(n-i)X(n-j)X(n-k)X(n-l),$$

converge in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ alla variabile aleatoria

$$Y(n)Y(n+m_1)Y(n+h+m_1)Y(n+h+m_1+m_2).$$

Quindi

$$\mathbb{E}\{Y(n)Y(n+m_1)Y(n+h+m_1)Y(n+h+m_1+m_2)\} =$$

$$= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \psi_i \psi_{j+m_1} \psi_{k+h+m_1} \psi_{l+h+m_1+m_2} \mathbb{E}\{X(n-i)X(n-j)X(n-k)X(n-l)\},$$

che può essere riscritta, grazie a (6.12), come

$$\begin{aligned} & \mathbb{E}\{Y(n)Y(n+m_1)Y(n+h+m_1)Y(n+h+m_1+m_2)\} = \\ & = (\theta-3) \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+m_1} \psi_{i+h+m_1} \psi_{i+h+m_1+m_2} + \Gamma(m_1)\Gamma(m_2) + \Gamma(h+m_1)\Gamma(h+m_2) + \\ & \quad + \Gamma(h+m_1+m_2)\Gamma(h+m_1+m_2)\Gamma(h). \end{aligned} \tag{6.13}$$

Calcoliamo il momento secondo incrociato delle variabili $Z_n(m)$, fissato n :

$$\begin{aligned} \mathbb{E}\{Z_n(m_1)Z_n(m_2)\} &= \frac{1}{(n+1)^2} \mathbb{E}\left\{\sum_{s=0}^n \sum_{t=0}^n Y(t)Y(t+m_1)Y(s)Y(s+m_2)\right\} = \\ &= \frac{1}{(n+1)^2} \sum_{s=1}^n \sum_{t=1}^n \mathbb{E}\{Y(t)Y(t+m_1)Y(s)Y(s+m_2)\}. \end{aligned}$$

Per (6.13), si ha

$$\begin{aligned} & \mathbb{E}\{Z_n(m_1)Z_n(m_2)\} = \\ & = \frac{1}{(n+1)^2} \sum_{s=0}^n \sum_{t=0}^n [\Gamma(m_1)\Gamma(m_2) + \Gamma(s-t)\Gamma(s-t+m_1+m_2) + \Gamma(s-t+m_2)\Gamma(s-t-m_1)] + \\ & \quad + \frac{1}{(n+1)^2} \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} (\theta-3)\sigma^4 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+m_1} \psi_{i+s-t} \psi_{i+s-t+m_2}. \end{aligned}$$

Siccome il filtro è assolutamente sommabile, converge anche la serie

$$\sum_{i=-\infty}^{\infty} |\psi_i \psi_{i+m_1} \psi_{i+s-t} \psi_{i+s-t+m_2}|.$$

Possiamo quindi cambiare il parametro di somma: posto $y = s - t$, otteniamo

$$\begin{aligned} \text{cov}\{Z_n(m_1), Z_n(m_2)\} &= \mathbb{E}\{Z_n(m_1)Z_n(m_2)\} - \Gamma(m_1)\Gamma(m_2) = \\ &= \frac{1}{n+1} \sum_{|y| < n+1} \left(1 - \frac{|y|}{n+1}\right) T_y \end{aligned}$$

dove

$$T_y = \Gamma(y)\Gamma(y-m_1+m_2) + \Gamma(y+m_2)\Gamma(y-m_1) + (\theta-3)\sigma^4 \sum_{i=-\infty}^{\infty} \psi_i \psi_{i+p} \psi_{i+y} \psi_{i+y+m_2}.$$

Poiché per ipotesi $\sum_{k=-\infty}^{\infty} |\Gamma(k)| < \infty$ e $\sum_{k=-\infty}^{\infty} |\psi_k| < \infty$, vale anche

$$\sum_{y=-\infty}^{\infty} |T_y| < \infty,$$

dunque abbiamo che

$$\lim_n (n+1) \text{cov}\{Z_n(m_1), Z_n(m_2)\} = \sum_{y=-\infty}^{\infty} \Gamma_y;$$

per la definizione di T_Y , dunque

$$\begin{aligned} & \lim_n (n+1) \text{cov} Z_n(m_1), Z_n(m_2) = \\ & = (\theta-3)\Gamma(m_1)\Gamma(m_2) + \sum_{j=-\infty}^{\infty} [\Gamma(j)\Gamma(j-m_1+m_2) + \Gamma(j+m_1)\Gamma(j-m_2)]. \end{aligned}$$

□

Supponiamo, momentaneamente, che il filtro ψ_j abbia un numero finito di termini non nulli e ricaviamo in questo caso la distribuzione asintotica delle variabili $Z_n(m)$. Sono necessari i teoremi di Hoeffding e Robbins e di Cramer Wold, che ci limitiamo ad enunciare (per una dimostrazione vedi [35]).

Teorema 6.3.2. *Supponiamo che $\{Z(n)\}_{n \in \mathbb{Z}}$ sia un processo stocastico stazionario in senso stretto centrato, con funzione di covarianza Ξ e m -dipendente (cioè se $n_1, n_2 \in \mathbb{Z}$ sono tali che $|n_2 - n_1| > m$ allora le variabili aleatorie $Z(n_1)$ e $Z(n_2)$ sono indipendenti). Valgono le seguenti affermazioni:*

1. $\lim_n \text{Var}\{\bar{Z}_n\} = v_m$

2. $\sqrt{n} \bar{Z}_n \xrightarrow{\mathcal{L}} Z$

dove $Z \sim N(0, v_m)$, $v_m = \Gamma(0) + 2 \sum_{k=1}^m \Gamma(k)$ e \bar{Z}_n indica la variabile aleatoria

$$\bar{Z}_n = \frac{1}{n+1} \sum_{i=0}^n Z(i).$$

Teorema 6.3.3. *Supponiamo che $\{\mathbf{X}_n\}_{n \in \mathbb{N}}$ sia una successione di vettori aleatori, a valori in \mathbb{R}^n , definiti su $(\Omega, \mathcal{F}, \mathbb{P})$. Allora*

$$\mathbf{X}_n \xrightarrow{\mathcal{L}} \mathbf{X}$$

se e solo se, per ogni $\lambda \in \mathbb{R}^n$, si ha

$$\langle \mathbf{X}_n, \lambda \rangle \xrightarrow{\mathcal{L}} \langle \mathbf{X}, \lambda \rangle.$$

Proposizione 6.3.3. *Supponiamo che esista $m \in \mathbb{N}$ tale che $\psi_j = 0$ per ogni j con $|j| > m$. Indichiamo con $Z_n(m)$ le variabili aleatorie*

$$Z_n(m) = \frac{1}{n+1} \sum_{i=0}^n Y(i)Y(i+m),$$

con \mathbf{Z}_n il vettore aleatorio $h+1$ -dimensionale

$$\mathbf{Z}_n = (Z_n(0), \dots, Z_n(h))^t.$$

Allora vale

$$\sqrt{n+1}(\mathbf{Z}_n - (\Gamma(0), \dots, \Gamma(m))^t) \xrightarrow{\mathcal{L}} \mathbf{N}_h(0, V),$$

dove $V = \{v_{i,j}\}_{i,j=0,\dots,h}$ è la matrice definita da

$$v_{i,j} = (\theta - 3)\Gamma(i)\Gamma(j) + \sum_{k=-\infty}^{\infty} [\Gamma(k)\Gamma(k-i+j) + \Gamma(k+i)\Gamma(k-j)].$$

Dimostrazione. Poiché la successione ψ_j contiene solo un numero finito di termini non nulli, vale

$$Y(n) = \sum_{j=-\infty}^{\infty} \psi_j X(n-j) = \sum_{j=-m}^m \psi_j X(n-j).$$

Il processo $\{X(n)\}_{n \in \mathbb{Z}}$ è a variabili indipendenti, dunque se $n_1, n_2 \in \mathbb{N}$ sono tali che $|n_1 - n_2| > 2m$, allora le variabili aleatorie $Y(n_1)$ e $Y(n_2)$ sono indipendenti.

Consideriamo i vettori aleatori

$$\mathbf{Y}(i) = (Y(i)Y(i), \dots, Y(i)Y(i+h))^t.$$

Poiché il processo $\{Y(n)\}_{n \in \mathbb{N}}$ è $2m$ -dipendente, la successione di vettori aleatori $\mathbf{Y}(i)$ è $2m + h$ -dipendente. Per la definizione di \mathbf{Z}_n , vale

$$\mathbf{Z}_n = \frac{1}{n+1} \sum_{i=0}^n \mathbf{Y}(i).$$

Dobbiamo dimostrare la convergenza in legge nel caso vettoriale: per il teorema di Cramer Wold ci basta dimostrare che per ogni $\lambda \in \mathbb{R}^{h+1}$ vale la convergenza

$$\sqrt{n+1}[\lambda^t \cdot \mathbf{Z}_n - (\Gamma(0), \dots, \Gamma(h))^t] \xrightarrow{\mathcal{L}} N(0, \lambda^t V \lambda).$$

La successione di vettori $\lambda^t \mathbf{Y}(i)$ è $2m + h$ -dipendente e possiamo applicare il teorema 6.3.2: dunque vale

$$\sqrt{n+1}[\lambda^t \cdot \mathbf{Z}_n - (\Gamma(0), \dots, \Gamma(h))^t] \xrightarrow{\mathcal{L}} N(0, v_m).$$

È semplice, ma un po' laborioso, constatare che $v_m = \lambda^t V \lambda$. Per il teorema di Cramer Wold dunque

$$\sqrt{n}(\mathbf{Z}_n - (\Gamma(0), \dots, \Gamma(m))) \xrightarrow{\mathcal{L}} N_{h+1}(0, V).$$

□

Dimostriamo che il risultato del lemma 6.3.3 può essere esteso anche a filtri lineari che non contengono solo un numero finito di termini non nulli. Per farlo ci serve un altro teorema, che ci limitiamo ad enunciare (per una dimostrazione, vedi [35].)

Teorema 6.3.4. *Supponiamo che X_n sia una successione di vettori aleatori, definiti su $(\Omega, \mathcal{F}, \mathbb{P})$ e a valori in \mathbb{R}^k . Supponiamo poi che per ogni $n \in \mathbb{N}$, Y_{n_j} sia una successione di vettori aleatori a valori in \mathbb{R}^k e definita su $(\Omega, \mathcal{F}, \mathbb{P})$, tale che:*

- Per ogni $j \in \mathbb{N}$ si abbia la convergenza per $n \rightarrow \infty$

$$Y_{n_j} \xrightarrow{\mathcal{L}} Y_j$$

- Per $j \rightarrow \infty$ valga

$$Y_j \xrightarrow{\mathcal{L}} Y.$$

- $\lim_{j \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\{|X_n - Y_n| > \epsilon\} = 0$ per ogni $\epsilon > 0$

allora

$$X_n \xrightarrow{\mathcal{L}} Y.$$

Proposizione 6.3.4. Consideriamo il processo $\{Y(n)\}_{n \in \mathbb{N}}$, $h \in \mathbb{R}$ e le successioni di vettori aleatori $\mathbf{Z}_n = (Z(0), \dots, Z(h+1))^t \in \mathbb{R}$. Allora

$$\sqrt{n+1}(\mathbf{Z}_n - (\Gamma(0), \dots, \Gamma(m))^t) \xrightarrow{\mathcal{L}} N_h(0, V),$$

dove $V = \{v_{i,j}\}_{i,j=0,\dots,h}$ è la matrice definita da

$$v_{i,j} = (\theta - 3)\Gamma(i)\Gamma(j) + \sum_{k=-\infty}^{\infty} [\Gamma(k)\Gamma(k-i+j) + \Gamma(k+i)\Gamma(k-j)].$$

Dimostrazione. Dato $m \in \mathbb{N}$ consideriamo il processo $\{Y^m(n)\}_{n \in \mathbb{N}}$ definito da

$$Y^m(n) = \sum_{j=-m}^m \psi_j X(n-j).$$

Fissato $n \in \mathbb{N}$, per $m \rightarrow \infty$ vale la convergenza

$$Y^m(n) \xrightarrow{L^2} Y(n).$$

Definiamo le variabili aleatorie

$$Z_n^m(k) = \frac{1}{n} \sum_{t=1}^n Y^m(t)Y^m(t+h),$$

per il lemma 6.3.3 si ha, indicando con Γ^m la funzione di covarianza del processo Y^m

$$\sqrt{n+1}(Z_n^m(0) - \Gamma^m(0), \dots, Z_n^m(h) - \Gamma^m(h)) \xrightarrow{\mathcal{L}} N(0, V_m),$$

dove $V_m = \{v_{i,j}^m\}_{i,j=0,\dots,h}$, è definita da

$$\{V_m\}_{i,j} = (\theta - 3)\Gamma_m(i)\Gamma_m(j) - \sum_{k=-\infty}^{\infty} \Gamma_m(k)\Gamma_m(k-i+j) + \Gamma_m(k+j)\Gamma_m(k-i).$$

Siccome

$$\Gamma_m(h) = \langle Y_m(n), Y_m(n+h) \rangle,$$

dove $\langle \cdot \rangle$ indica il prodotto scalare in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, e si ha che in $L^2(\Omega, \mathcal{F}, \mathbb{P})$

$$Y_m(n) \rightarrow Y(n) \quad Y_m(n+h) \rightarrow Y(n+h),$$

la successione di funzioni $\{\Gamma^m\}_{m \in \mathbb{N}}$ converge puntualmente a Γ . Dunque anche i termini della matrice V_m convergono puntualmente a V e dunque abbiamo che la successione di variabili aleatorie definita da

$$\{\mathbf{T}_n\}_{n \in \mathbb{N}}$$

dove $\mathbf{T}_i \sim N_{h+1}(0, V_m)$ converge il legge alla variabile aleatoria $\mathbf{Y} \sim N(0, V)$, dunque per il teorema 6.3.4 si conclude. \square

6.4 Inferenza statistica sulla funzione di covarianza nel caso il processo sia di Markov

Grazie a quanto detto in 3.3 sappiamo che il processo $\{Y(n)\}_{n \in \mathbb{N}}$ è di Markov se e solo se

$$\Gamma(n) = \Gamma(0)\theta^n$$

dove $\rho(i)$ è definito da

$$\theta = \frac{\Gamma(1)}{\Gamma(0)}.$$

In questo paragrafo vogliamo utilizzare i correlogrammi per costruire uno stimatore di θ , di cui tenteremo di dimostrare delle proprietà asintotiche. Consideriamo la successione di variabili aleatorie

$$\frac{\hat{\Gamma}_n(1)}{\hat{\Gamma}_n(0)} \tag{6.14}$$

Continuiamo ad ipotizzare che valgano le ipotesi specificate all'inizio del paragrafo precedente. La successione di variabili aleatorie è ben definita e si può dimostrare che si tratta di una successione di stimatori asintoticamente corretta di θ .

Sappiamo infatti, per la proposizione 6.3.4, che il vettore

$$(\hat{\Gamma}_n(0), \hat{\Gamma}_n(1))$$

è asintoticamente normale con vettore delle medie $(\Xi(0), \Xi(1))$ e matrice di covarianza V , descritta nella proposizione a cui facciamo riferimento. Consideriamo la funzione

$$g : A \rightarrow \mathbb{R}$$

$$(x, y) \mapsto \frac{x}{y}$$

dove

$$A := \{(x, y) \in \mathbb{R}^2 | x \neq 0\}.$$

Essa è continua sull'insieme di definizione; notiamo poi che l'insieme A^c ha misura nulla secondo ogni misura equivalente alla misura di Lebesgue bi-dimensionale. La seguente proposizione consente di determinare la distribuzione asintotica degli stimatori.

Proposizione 6.4.1. *Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e $\{X_n\}_{n \in \mathbb{N}}$ una successione di vettori aleatori k -dimensionali definiti su di esso. Supponiamo che*

$$\frac{1}{c_n} \cdot (X_n - \mu) \xrightarrow{\mathcal{L}} Y$$

dove $Y \sim N_k(0, \Sigma)$ e la successione $\{c_n\}_{n \in \mathbb{N}}$ è infinitesima. Sia inoltre

$$g = (g_1, \dots, g_k)$$

una funzione con derivate parziali continue nel punto μ , con $D = \frac{dg}{dx}(\mu)$, e supponiamo che la matrice

$$D\Sigma D^t$$

abbia tutti gli elementi sulla diagonale principale diversi da 0, allora

$$\frac{1}{c_n} \cdot (g(X_n) - g(\mu)) \xrightarrow{\mathcal{L}} Z$$

dove $Z \sim N(0, D\Sigma D^t)$.

Dimostrazione. Consideriamo $j \in \{1, \dots, k\}$ e la successione di variabili aleatorie $\{X_{n,j}\}_{n \in \mathbb{N}}$. Per il teorema di Cramer Wold è facile concludere che la successione

$$\frac{X_{n,j} - \mu_j}{c_n} \xrightarrow{\mathcal{L}} X, \tag{6.15}$$

dove $X \sim N(\mu_j, \sigma_{j,j})$. Siccome la successione delle variabili in 6.15 è convergente in legge, essa è una famiglia tesa di variabili aleatorie, dunque

$$X_{n,j} = \mu_j + O_{\mathbb{P}}(c_n).$$

Ora, siccome per ipotesi g ha derivate parziali continue in μ , possiamo applicare a $g(X_n)$ lo sviluppo di Taylor in probabilità fino al primo termine, ottenendo:

$$g(X_n) = g(\mu) + D(X_n - \mu) + o_{\mathbb{P}}(c_n),$$

dunque che

$$\frac{g(X_n) - g(\mu)}{c_n} = \frac{D(X_n - \mu)}{c_n} + o_{\mathbb{P}}(1).$$

Per quanto detto precedentemente abbiamo che

$$\frac{D(X_n - \mu)}{c_n} \xrightarrow{\mathcal{L}} F$$

dove il vettore aleatorio F è gaussiano con matrice di covarianza $D\Sigma D^t$, il teorema di Slutsky permette di concludere. \square

Torniamo alla distribuzione asintotica degli stimatori: calcoliamo il gradiente di g in $(\Gamma(0), \Gamma(1))$

$$\nabla g(\mu) = J(g)(\Gamma(0), \Gamma(1)) = \left(\frac{\Gamma(1)}{\Gamma^2(0)}, \frac{1}{\Gamma(0)} \right)$$

Dunque, applicando la proposizione precedente, sappiamo che la successione 6.14 è asintoticamente corretta e normale, con varianza $\nabla g(\mu) V \nabla g(\mu)^t$. Ricordando come è stata definita la matrice V , ed indicando con $\rho(t)$ la funzione $\frac{\Gamma(t)}{\Gamma(0)}$:

$$\nabla g(\mu) V \nabla g^t = \sum_{k=-\infty}^{k=\infty} [2\rho(k)^2 + 2\rho(k+1)\rho(k-1) - 4\rho(1)\rho(k)^2].$$

Siccome la funzione di covarianza è data da

$$\Xi(t) = \Xi(0)\theta^{|t|},$$

abbiamo

$$\nabla g(\mu)V\nabla g(\mu)^t = \sum_{k=-\infty}^{k=\infty} [2\theta^{2|k|} + \theta^{|k-1|}\theta^{|k+1|} + 2\theta^2\theta^{2|k|}].$$

Semplificando risulterà dunque:

$$\nabla g(\mu)V\nabla g(\mu)^t = \frac{7\theta^2 + 4 - \theta^4}{1 - \theta^2},$$

quindi la successione di stimatori 6.14 verifica

$$\sqrt{n} \cdot \left(\frac{\hat{\Gamma}_n(1)}{\hat{\Gamma}_n(0)} - \theta \right) \xrightarrow{\mathcal{L}} N\left(0, \frac{7\theta^2 + 4 - \theta^4}{1 - \theta^2}\right).$$

6.5 Risultati ottenuti

Nella tabella 6.5 riportiamo le stime della media μ , di $\Gamma(0)$ e di θ ottenute con gli stimatori descritti all'interno del capitolo.

μ	$\Gamma(0)$	θ
12.492	0.888	0.191

Tabella 6.1: Stime ottenute con gli stimatori descritti.

Vi sono varie ragioni che inducono a preferire le stime date nel capitolo 5 a queste: prima di tutto non abbiamo testato in alcun modo le ipotesi fatte sulla struttura o sulla distribuzione spettrale del processo. Anche ammesso che esse siano realizzate, poi, i risultati che abbiamo dimostrato hanno valore solo asintotico e noi abbiamo solo 45 osservazioni del processo.

Capitolo 7

Tentativo di convalida del modello

In questo capitolo cominciamo a considerare le misure relative agli individui sieropositivi. Il modello proposto da [7] è il seguente.

Supponiamo che il processo stocastico che modella il decadimento della concentrazione di linfociti $CD4^+$ per mm^3 di sangue di individui sieropositivi sia $\{X(t)\}_{t \in \mathbb{R}^+}$, dove $t = 0$ rappresenta il momento dell'infezione da HIV. In [7] l'autore suppone che, una volta applicata la trasformazione logaritmica, il processo

$$\{\log X(t)\}_{t \in \mathbb{R}^+};$$

sia un processo gaussiano, con struttura del second'ordine stazionaria: è dunque ben definita una funzione di covarianza. Egli ipotizza inoltre che quest'ultima sia uguale a quella che è già stata descritta e stimata nel paragrafo precedente. La funzione delle medie del processo proposta, almeno inizialmente, è:

$$\mathbb{E}\{\log X(t)\} = \mu - \delta t$$

dove μ è la media che è già stata stimata nel capitolo precedente, mentre δ è un parametro che deve essere stimato.

Risulta necessario fin d'ora apportare alcune modifiche al modello, affinché esso risulti compatibile con l'analisi svolta fin'ora. Nel capitolo 4 abbiamo discusso l'opportunità di adottare una trasformazione diversa dal logaritmo ed abbiamo scelto la funzione

$$f_\lambda = \frac{x^\lambda - 1}{\lambda} \quad \lambda = 0.2.$$

In questo capitolo essenzialmente verrà testato il modello su due livelli, utilizzando le misure dei sieroconvertiti, descritte nel paragrafo seguente.

Dapprima sarà verificato che la trasformazione individuata sia compatibile con il modello, ossia che $\{f_\lambda X(t)\}_{t \in \mathbb{R}^+}$ sia un processo gaussiano. Questo sarà mostrato nel paragrafo 7.2.

In secondo luogo, nel paragrafo 7.3 tenteremo di verificare che la funzione della media del processo è lineare: questo porterà ad individuare dei possibili errori nel modello adottato.

7.1 I sieroconvertiti: potenzialità e problemi dei dati

Il tempo $t = 0$ corrisponde dunque al momento dell'infezione. Salvo casi particolari, il tempo trascorso dall'infezione alla scoperta della sieropositività è ignoto.

Nel testo citato l'autore suppone che tale periodo sia rappresentato da una variabile aleatoria T , indipendente dal processo $\{X(t)\}_{t \in \mathbb{R}^+}$. Negli individui che fin dall'inizio del monitoraggio sono sieropositivi, dunque, le misure a nostra disposizione non sono realizzazioni dirette del processo $\{X(t)\}_{t \in \mathbb{R}^+}$, ma del processo $\{X(T + t)\}_{t \in \mathbb{R}^+}$.

Fra i dati di cui disponiamo sono presenti anche quelli relativi ad individui sieroconvertiti, ossia individui passati dallo stato di sieronegatività allo stato di sieropositività durante il monitoraggio. Questi dati potenzialmente sarebbero molto importanti perché, se il tempo d'infezione è approssimato a quello della prima visita in cui essi risultano sieropositivi, consentirebbero di osservare direttamente delle realizzazioni del processo $\{X(t)\}_{t \in \mathbb{R}^+}$.

Si pongono però diversi problemi all'analisi di questi dati. Il primo riguarda l'approssimazione del tempo $t = 0$, con un range di un semestre. Un secondo problema è dovuto alla mancanza di informazioni circa le cure a cui i pazienti sono stati sottoposti durante il monitoraggio. Nel database che utilizziamo, infatti, sono presenti delle indicazioni riguardanti i tipi di farmaci assunti, ma riguardano solo le visite successive alla numero 13, che corrisponde all'anno 1990. L'anno di commercializzazione del primo farmaco antivirale impiegato nella terapia dei pazienti affetti da HIV, la Zidovudina o AZT, è invece il 1987. Secondo [51] già nel marzo del 1987, 4500 pazienti affetti da AIDS (un terzo di tutti gli americani viventi affetti da AIDS) ricevettero tale farmaco.

Gli autori di [23] inoltre sostengono che esso abbia un effetto sul decadimento di linfociti $CD4+$; per studiare la naturale evoluzione del fenomeno è quindi necessario scartare preventivamente tutti i dati posteriori al 1987. Questo diminuisce drasticamente i dati a nostra disposizione: risultano al massimo 5 visite relative ai singoli pazienti, a partire dalla prima in cui essi sono risultati positivi. Di seguito riportiamo un grafico che riassume i dati relativi ai sieroconvertiti.

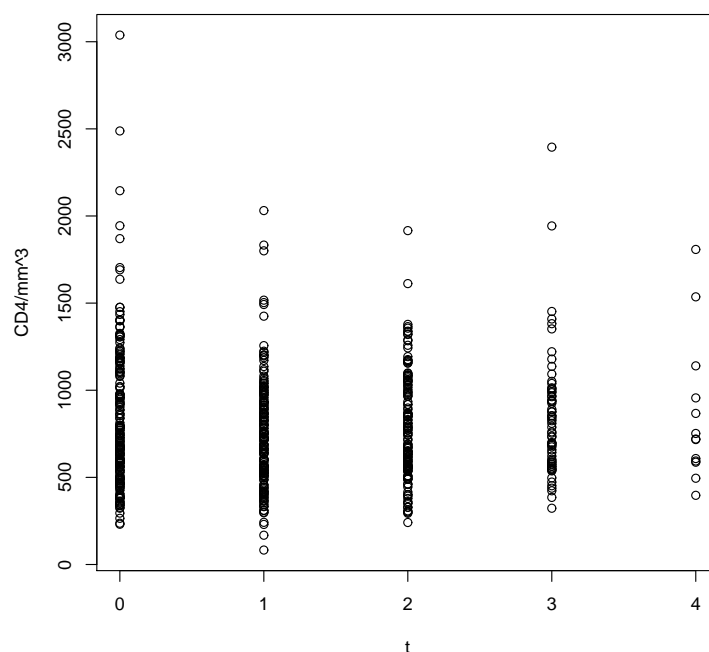


Figura 7.1: Il grafico riporta il numero di linfociti $CD4+$ per mm^3 di sangue nei sieroconvertiti in funzione dei tempi trascorsi dall'infezione, approssimati come detto sopra. Si può notare come la taglia dei campioni decresca all'aumentare di t .

7.2 La verifica della trasformazione

Basandoci sulle misurazioni relative ai sieroconvertiti, abbiamo a disposizione le realizzazioni di un campione a variabili indipendenti delle variabili $X(i)$ per $i = 1, \dots, 5$. La prima cosa di cui ci accertiamo è che la trasformazione f_λ , determinata nel capitolo 4, sia compatibile con il modello, ossia che il processo

$$\{f_\lambda(X(t))\}_{t \in \mathbb{R}}$$

sia gaussiano. Noi verifichiamo parzialmente questa ipotesi, testando l'ipotesi di gaussianità separatamente sulle osservazioni delle variabili $f_\lambda(X(0)), \dots, f_\lambda(X(4))$. Utilizziamo gli stessi test usati nel capitolo 4; i risultati sono riportati nella tabella 7.2.

	Anderson-Darling	Cramer-Von-Mises	Pearson	Kolmogorov-Smirnov
$t = 0$	0.074	0.065	0.161	0.107
$t = 1$	0.105	0.121	0.336	0.150
$t = 2$	0.325	0.392	0.538	0.150
$t = 3$	0.359	0.478	0.537	0.643
$t = 4$	0.421	0.378	0.727	0.397

Tabella 7.1: Stime ottenute con gli stimatori descritti.

Con un livello di significatività di 0.05, nessuno dei test respinge l'ipotesi di gaussianità, per ognuna delle cinque visite. Osserviamo, per inciso, che non è possibile confrontare i *p-values* relativi ad uno stesso test per tempi differenti, dal momento che i campioni considerati hanno taglie diverse.

7.3 Verifica della linearità del processo

Supponiamo che il processo

$$\{Y(t)\}_{t \in \mathbb{R}^+} = \{f_\lambda(X(t))\}_{t \in \mathbb{R}^+}$$

sia gaussiano. Esso può essere scomposto in una parte deterministica ed in un processo di errore

$$Y(t) = \mu(t) + \epsilon(t)$$

In questo paragrafo vogliamo tentare di verificare che la parte deterministica sia una funzione lineare di t , ossia che

$$\mu(t) = a + bt$$

per determinati parametri a e b . Se questo fosse verosimile, potremmo dare una prima stima di a e b , dunque una prima stima di δ , a partire dalle misurazioni riguardanti i sieroconvertiti. Continueremo a supporre per il resto del capitolo, in accordo con [7], che la funzione di covarianza del processo sia uguale a quella dell'analogo processo nei sani, descritta nel capitolo precedente.

In questo paragrafo supporremo inoltre che la componente deterministica sia ben approssimata da una funzione polinomiale in t , con grado minore o uguale a 3: considerando le variabili aleatorie $Y(0), \dots, Y(4)$ si presenta perciò la situazione di un modello lineare generalizzato¹.

7.3.1 Il modello lineare generalizzato e lo stimatore di Aitken

Consideriamo k coppie di variabili aleatorie $(X_i, Y_i)_{i=1, \dots, k}$, un vettore aleatorio centrato $(\epsilon_1, \dots, \epsilon_k)$ e il modello di regressione

$$Y = \mathbf{X}\beta + \epsilon \tag{7.1}$$

Dove $\beta \in \mathbb{R}^p$ e \mathbf{X} è la cosiddetta *design matrix*, definita da

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 & \dots & X_1^p \\ 1 & X_2 & \dots & X_2^p \\ \dots & \dots & \dots & \dots \\ 1 & X_k & \dots & X_k^p \end{pmatrix}$$

Consideriamo la *design matrix* come non aleatoria; supponiamo che essa abbia rango massimo e che la matrice di covarianza W del vettore aleatorio ϵ sia

¹Il termine *modello lineare generalizzato* ha significati diversi: alcuni, come [10] lo intendono con la nostra definizione, altri, come [50], intendono una situazione diversa, in cui $Y = f(X) + \epsilon$, ma f non è direttamente polinomiale.

definita positiva. In questo caso il modello di regressione 7.1 si dice *modello di regressione generalizzato* o *modello lineare generalizzato*. Lo scopo di questo paragrafo è generalizzare ai modelli lineari generalizzati i noti risultati che valgono nei modelli di regressione standard.

Notiamo che, applicando opportune trasformazioni lineari ad ambo i membri dell'equazione 7.1, è possibile ricondursi al caso di un modello di regressione standard. W è definita positiva, quindi anche W^{-1} lo è: sono ben definite le radici delle matrici $M = \sqrt{W}$ e $N = \sqrt{W^{-1}}$. Visto che W^{-1} è l'inversa di W , si ha

$$NMMN = NWN = I.$$

Dunque, applicando la trasformazione lineare definita dalla matrice N ad ambo i membri di 7.1, si ottiene:

$$NY = N\mathbf{X}\beta + N\epsilon;$$

Detti $Y' = NY$, $X' = NX$ e $\epsilon' = N\epsilon$, il nuovo vettore di errore ϵ' è ancora centrato con matrice di covarianza $NWN^t = NWN = I$; inoltre la nuova *design matrix* $N\mathbf{X}$ è ancora di rango massimo, perché N è non singolare. Un modello di regressione in cui la *design matrix* abbia rango massimo e il vettore degli errori sia costituito da variabili gaussiane indipendenti ed equidistribuite si dice solitamente *modello di regressione standard*. Dunque il modello

$$Y' = \mathbf{X}'\beta + \epsilon' \tag{7.2}$$

è un modello di regressione standard.

Il teorema di Gauss-Markov-Aitken dimostra l'esistenza, nel caso di un modello lineare generalizzato, di uno stimatore ottimale nella classe degli stimatori lineari corretti di β .

Proposizione 7.3.1 (Gauss-Markov-Aitken). *Supponiamo che*

$$Y = \mathbf{X}\beta + \epsilon$$

sia un modello lineare generalizzato, con vettore di errore omoschedastico, ossia, detta σ^2 la varianza comune, si abbia

$$W = \sigma^2 W'$$

con W' matrice definita positiva e con diagonale principale unitaria. Lo stimatore

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

è uno stimatore ottimale fra gli stimatori lineari e corretti di β .

Dimostrazione. Notiamo che $\hat{\beta}$ corrisponde allo stimatore dei minimi quadrati del modello (7.2), infatti

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}^t W^{-1} \mathbf{X})^{-1} \mathbf{X} W^{-1} Y = (\mathbf{X}^t N \cdot N \mathbf{X})^{-1} \mathbf{X} N \cdot N Y = \\ &= (\mathbf{X}^t \mathbf{X}')^{-1} \mathbf{X}' Y'.\end{aligned}$$

Per il teorema di Gauss Markov esso è uno stimatore lineare e corretto di β . Vediamo che è anche ottimale (rispetto al costo quadratico medio), nella classe degli stimatori lineari corretti. Consideriamo per questo un altro stimatore lineare

$$\check{\beta} = \check{C} Y,$$

ed indichiamo con \hat{C} la matrice $(\mathbf{X}^t W^{-1} \mathbf{X})^{-1} \mathbf{X} W^{-1}$. Definendo $D = \check{C} - \hat{C}$, possiamo scrivere

$$\check{\beta} = (\hat{C} + D) Y.$$

Ora imponiamo la condizione di correttezza: si ottiene

$$\mathbb{E}\{\check{\beta}\} = \hat{C} \mathbf{X} \beta + D \mathbf{X} \beta = \beta + D \mathbf{X} \beta,$$

da cui si deve avere $D \mathbf{X} \beta = 0$ e quindi, siccome $\beta \in \mathbb{R}^p$ è un generico vettore, $D \mathbf{X} = 0$. Valutiamo la matrice di covarianza dello stimatore, ricordando che esso è corretto otteniamo che:

$$\text{Var}\{\check{\beta}\} = \mathbb{E}\{\check{C} \epsilon \epsilon^t \check{C}^t\} = \sigma^2 (\hat{C} + D) W (\hat{C}^t + D^t).$$

La condizione $D \mathbf{X} = 0$ implica che

$$\hat{C} W D^t = 0 \quad \text{e} \quad D W \hat{C}^t = 0.$$

dunque

$$\sigma^2 (\hat{C} + D) W (\hat{C}^t + D^t) = \sigma^2 \hat{C} W \hat{C}^t + \sigma^2 D W D^t = \text{Var}\hat{\beta} + \sigma^2 D W D^t,$$

perché per il teorema di Gauss Markov $\text{Var}\{\hat{\beta}\} = \sigma^2 \hat{C} W \hat{C}^t$. La matrice $D W D^t$ è definita positiva, perché W lo è, quindi per ogni $x \in \mathbb{R}^p$ si ha:

$$\text{Var}\{x^t \check{\beta}\} \geq \text{Var}\{x^t \hat{\beta}\},$$

il che implica

$$MSE(\check{\beta}) \geq MSE(\hat{\beta}),$$

dove MSE indica il costo quadratico medio. □

7.3.2 Distorsione e varianza: il principio di parsimonia

Supponiamo di avere un modello lineare generalizzato

$$Y = \mathbf{X}\beta + \epsilon,$$

dove $Y = (Y_1, \dots, Y_n)$ e $X = (X_1, \dots, X_n)$ e $\beta \in \mathbb{R}^p$.

Per quanto detto nel paragrafo 7.3.1, se supponiamo che la matrice di covarianza W degli errori sia nota, conosciamo uno stimatore con delle proprietà ottimali. In questo paragrafo ci occupiamo del problema preliminare: come si sceglie un modello di regressione? Che rischi si corrono sottostimando o sovrastimando il numero di variabili di regressione?

Per semplicità supporremo che la matrice W sia diagonale e definita positiva. In questa situazione, lo stimatore di Aitken è detto anche dei minimi quadrati pesati e sarà:

$$\hat{\beta} = (\mathbf{X}'W^{-1}\mathbf{X})^{-1}\mathbf{X}'W^{-1}Y.$$

Supponiamo che il modello di regressione corretto sia:

$$Y = \mathbf{X}_0\beta_0 + \epsilon_0$$

con $\beta_0 \in \mathbb{R}^p$. Sia $q > p$ e consideriamo il modello di regressione sovrastimato

$$Y = \mathbf{X}\beta + \epsilon \quad \text{SOVRASTIMATO}$$

verificato per il vettore $\beta = (\beta_0, \mathbf{0}_{q-p})$.

Lo stimatore di Aitken, nel modello sovrastimato, sarà dato da:

$$\hat{\beta} = (\mathbf{X}'W^{-1}\mathbf{X})^{-1}\mathbf{X}'W^{-1}Y$$

Studiamone valore atteso e matrice di covarianza. Per il teorema di Gauss-Markov-Aitken si tratta ancora di uno stimatore corretto del vettore β e dunque, denotata con π la proiezione di \mathbb{R}^q sul sottospazio generato dai primi p vettori della base canonica, abbiamo che $\pi(\mathbb{E}\{\hat{\beta}\}) = \beta_0$.

Valutiamo la matrice di covarianza dello stimatore. Per lo stesso teorema, la matrice di covarianza di $\hat{\beta}$ è data da $(\mathbf{X}'W^{-1}\mathbf{X})^{-1}$. La matrice di regressione può essere decomposta in due blocchi di colonne, una di lunghezza p , l'altra di lunghezza $q - p$

$$\mathbf{X} = [\mathbf{X}_0, \mathbf{X}_1]$$

Dunque, una volta definita

$$C = [\mathbf{0}_p, \mathbf{I}_{q-p}]$$

e interpretando lo stimatore $\hat{\beta}_0$ come una variabile aleatoria a valori in \mathbb{R}^q (formalmente dovremmo scrivere $(\mathbf{0}_p, \hat{\beta}_0)$), la matrice di covarianza di $\hat{\beta}_0$ è data da

$$(\mathbf{X}^t W^{-1} \mathbf{X})^{-1} - (\mathbf{X}^t W^{-1} \mathbf{X})^{-1} C^t [C (\mathbf{X}^t W^{-1} \mathbf{X})^{-1} C^t]^{-1} C (\mathbf{X}^t W^{-1} \mathbf{X})^{-1}.$$

La differenza fra le due matrici di covarianza (avendo immerso lo stimatore $\hat{\beta}_0$ in \mathbb{R}^q esse hanno la stessa dimensione) è

$$\text{Var} \hat{\beta} - \text{Var} \hat{\beta}_0 = C^t (\mathbf{X}^t W^{-1} \mathbf{X})^{-1} [C (\mathbf{X}^t W^{-1} \mathbf{X})^{-1} C^t]^{-1} C (\mathbf{X}^t W^{-1} \mathbf{X})^{-1}.$$

Si verifica facilmente che tale matrice è semidefinita positiva, dunque che per ogni $x \in \mathbb{R}^p$

$$\text{Var}\{x^t \hat{\beta}\} \geq \text{Var}\{\pi(x)^t \hat{\beta}_0\},$$

quindi sovrastimare il numero di variabili di regressione non porta ad una distorsione dello stimatore, ma porta ad un incremento della varianza delle componenti dello stimatore, quindi ad un incremento del suo costo quadratico medio.

Vediamo cosa accade nel caso di sottostima delle variabili di regressione. Supponiamo perciò che il reale modello di regressione sia dato da

$$Y = \mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \epsilon$$

con $\beta_1 \in \mathbb{R}^p$ e $\beta_2 \in \mathbb{R}^q$. Ora supponiamo di sottostimare il numero di variabili di regressione, cioè di considerare il modello

$$Y = \mathbf{X}_1 \beta_1 + \epsilon \quad \text{SOTTOSTIMATO}$$

A meno che $\beta_2 = 0$, esso non è verificato. Lo stimatore di Aitken per β_1 ricavato dal modello sottostimato sarà:

$$\hat{\beta}_1 = (\mathbf{X}_1^t W^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1^t W Y$$

Esso non è uno stimatore corretto: abbiamo infatti

$$\mathbb{E}\{\hat{\beta}_1\} = (\mathbf{X}_1^t W^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1^t W (\mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2) =$$

$$= \beta_1 + (\mathbf{X}_1^t W^{-1} \mathbf{X}_1)^{-1} \mathbf{X}_1^t W \mathbf{X}_2 \beta_2 \neq \beta_1.$$

Con procedimento del tutto analogo a quello precedente possiamo dimostrare che la differenza fra la matrice di covarianza di $\hat{\beta}$ e quella di $\hat{\beta}_1$ è semidefinita positiva (previo considerare $\hat{\beta}_1$ come una variabile aleatoria con codominio di dimensione uguale a quello di $\hat{\beta}$).

Quanto detto può essere riassunto in questo modo: dato uno stimatore $\check{\beta}$, il costo quadratico medio è scomposto in

$$MSE(\check{\beta}) = \text{Var}(\check{\beta}) + \|\mathbb{E}\{\check{\beta}\} - \beta\|^2.$$

Il primo termine è la varianza dello stimatore (varianza è un termine improprio: nel caso vettoriale è la somma delle varianze delle componenti), mentre il secondo è la *distorsione*. Abbiamo visto che aumentando il numero di variabili di regressione, utilizzando i relativi stimatori di Aitken, la distorsione diminuisce, fino ad annullarsi se il numero di variabili è maggiore o uguale a quelle del modello reale. Viceversa la varianza dello stimatore cresce.

Il problema può essere descritto quindi in termini di conflitto fra varianza e distorsione. Vedremo nei paragrafi successivi che la scelta delle variabili si basa frequentemente sul *principio di parsimonia*, noto anche come *rasoio di Ockham*:

è inutile fare con più ciò che si può fare con meno.

7.3.3 La divergenza di Kullback-Leibler

Sia (Ω, \mathcal{F}) uno spazio di probabilità, in questo paragrafo introduciamo una *distanza* fra alcune misure di probabilità definite su di esso. Dimosteremo che in realtà questa funzione non possiede le proprietà richieste ad una funzione per essere una metrica: il termine *distanza* risulta quindi infelice; essa verrà indicata genericamente come *divergenza*.

Definizione 7.3.1. *Siano \mathbb{P} e \mathbb{Q} due misure di probabilità definite sullo spazio (Ω, \mathcal{F}) e supponiamo che $\mathbb{Q} \ll \mathbb{P}$, si dice divergenza di Kullback-Leibler da \mathbb{P} a \mathbb{Q} il valore (supposto che esso abbia senso)*

$$D_{KL}(\mathbb{P}||\mathbb{Q}) = - \int_{\Omega} \log \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P},$$

dove $\frac{d\mathbb{Q}}{d\mathbb{P}}$ indica la derivata di Radon-Nikodym.

Ovviamente, se entrambe le probabilità sono assolutamente continue rispetto ad una misura μ , dette p e q le relative densità, si ha:

$$D_{KL}(\mathbb{P}||\mathbb{Q}) = - \int_{\Omega} p \log \frac{q}{p} d\mu.$$

D_{KL} verifica la prima proprietà delle metriche: se \mathbb{P} e \mathbb{Q} sono tali che $D_{KL}(\mathbb{P}||\mathbb{Q})$ è ben definita, vale

$$D_{KL}(\mathbb{P}||\mathbb{Q}) \geq 0$$

e si ha l'uguaglianza se e solo se $\mathbb{Q} = \mathbb{P}$. Questo segue immediatamente dalla *disuguaglianza di Gibbs*, che proveremo ora.

Proposizione 7.3.2. *Siano \mathbb{P} e \mathbb{Q} due misure di probabilità definite sullo spazio (Ω, \mathcal{F}) . Supponiamo che $\mathbb{Q} \ll \mathbb{P}$ e che $\log \frac{d\mathbb{P}}{d\mathbb{Q}}$ sia \mathbb{P} integrabile, allora*

$$\int_{\Omega} \log \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{P} \leq 0,$$

e l'uguaglianza vale se e solo se $\mathbb{P} = \mathbb{Q}$.

Dimostrazione. Le variabili aleatorie in questione sono, quasi certamente, ben definite. Ricordando la disuguaglianza, valida per ogni $x > 0$

$$\log x \leq x - 1,$$

abbiamo che

$$\begin{aligned} \int_{\Omega} \log \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P} &\leq \int_{\Omega} \left(\frac{d\mathbb{Q}}{d\mathbb{P}} - 1 \right) d\mathbb{P} = \\ &= \int_{\Omega} d\mathbb{Q} - \int_{\Omega} d\mathbb{P} = 0. \end{aligned}$$

L'uguaglianza vale se e solo se $\log \frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{d\mathbb{Q}}{d\mathbb{P}} - 1$, almeno \mathbb{P} -quasi certamente: questo implica che una versione di $\frac{d\mathbb{Q}}{d\mathbb{P}}$ sia la costante 1, dunque che le due misure siano uguali. \square

Dimostriamo infine perché la divergenza di Kullback-Leibler non è una metrica. La funzione non è simmetrica: prima di tutto perché, se risulta ben definita $D_{KL}(\mathbb{P}||\mathbb{Q})$, non è detto che si possa dire altrettanto per $D_{KL}(\mathbb{Q}||\mathbb{P})$. Anche quando ciò accade, tuttavia, la simmetria non è verificata. Prendiamo per esempio due misure sullo spazio $((0, 1), \mathcal{B}(0, 1))$: \mathbb{P} sia la probabilità uniforme e \mathbb{Q} abbia densità, rispetto alla misura di Lebesgue ristretta all'intervallo, $2x$. Si verifica facilmente che $D_{KL}(\mathbb{P}||\mathbb{Q}) \neq D_{KL}(\mathbb{Q}||\mathbb{P})$.

7.3.4 Il criterio di informazione secondo Akaiake

Consideriamo un campione di variabili aleatorie (Y_1, \dots, Y_n) , ed un modello statistico definito dalla verosimiglianza, rispetto alla misura di Lebesgue n -dimensionale

$$L(\theta, x_1, \dots, x_n) \quad \text{con } \theta \in \Theta \subset \mathbb{R}^p.$$

Supponiamo che risulti definito uno stimatore di massima verosimiglianza per θ , che indicheremo con $\hat{\theta}$.

Si dice *criterio di informazione* una variabile aleatoria nella forma

$$IC_k = -2(\log(L(\hat{\theta}, Y_1, \dots, Y_n)) - \phi(n)k)$$

in cui k indica la dimensione di Θ e ϕ è una funzione in genere crescente. Un criterio di informazione può essere dunque scomposto in due parti: la prima valuta la distorsione del modello, mentre la seconda penalizza la sovrastima del numero parametri, che induce un aumento della varianza. È, in sostanza, una possibile matematizzazione del principio di parsimonia.

Come sarà chiaro in seguito, i criteri di informazione non hanno valore assoluto, cioè vengono utilizzati per confrontare quanto diversi modelli si adattino alle osservazioni: una volta scelto il migliore, però, non forniscono alcuna informazione diretta su quanto questo si adatti ai dati a nostra disposizione.

Uno dei più noti criteri di informazione è quello di Akaiake. Indicato con AIC_k , si ottiene ponendo $\phi(n) = 1$:

$$AIC_k = -2(\log(L(\hat{\theta}, Y_1, \dots, Y_n)) - k)$$

Il resto del paragrafo è finalizzato all'esposizione di come il criterio di informazione di Akaiake possa essere costruito basandosi sulla divergenza di Kullback-Leibler.

Noi considereremo solo il caso in cui (Y_1, \dots, Y_n) sia un campione di variabili aleatorie indipendenti ed identicamente distribuite. Il caso in cui poi lo applicheremo, in realtà, è quello di regressione, in cui le coppie $(X_i, Y_i), \dots, (X_n, Y_n)$ sono variabili che costituiscono un modello di regressione

$$Y = \mathbf{X}\beta + \epsilon.$$

La costruzione del criterio in questo caso è sostanzialmente analoga, ma è resa più complicata dalla teoria degli stimatori di massima verosimiglianza

in un modello di regressione, che è più complessa (per una trattazione dettagliata del caso di regressione, si veda [18]).

Siccome sfrutteremo risultati asintotici sugli stimatori di massima verosimiglianza, supponiamo di avere una successione di variabili aleatorie reali $\{Y_n\}_{n \in \mathbb{N}}$, indipendenti ed identicamente distribuite, la cui distribuzione abbia densità g rispetto alla misura di Lebesgue. Consideriamo poi una famiglia di probabilità di densità $f(x, \theta)$, con $\theta \in \Theta$. Per prima cosa si deve stabilire come valutare la differenza fra due misure di probabilità: in [2] l'autore propone di utilizzare la divergenza di Kullback-Leibler, ove essa risulti ben definita. Indicando le probabilità con le rispettive densità abbiamo

$$D_{KL}(g(x)||f(x, \theta)) = \int g(x) \log g(x) dx - \int g(x) \log f(x, \theta) dx.$$

Supponiamo che la funzione

$$\theta \mapsto \int g(x) \log g(x) dx - \int g(x) \log f(x, \theta) dx \quad (7.3)$$

sia differenziabile e che essa abbia un unico punto di minimo interno a Θ . Tale punto, indicato con $\bar{\theta}$, è detto il parametro *meno falso*. Notiamo che (7.3) è differenza di due funzioni, la prima delle quali è costante; il parametro meno falso è dunque un punto di massimo di

$$\theta \mapsto \int g(x) \log f(x, \theta) dx. \quad (7.4)$$

In sostanza Akaiake propone di quantificare la differenza esistente fra un modello statistico proposto, con verosimiglianza $f(x, \theta)$, e la reale funzione di densità $g(x)$ mediante la minima divergenza di Kullback-Leibler: siccome essa è differenza di due quantità, la prima delle quali dipende esclusivamente da $g(x)$, si può adottare come quantificazione il valore

$$R = \int g(x) \log f(x, \bar{\theta}) dx.$$

Ora però $g(x)$ è incognita, dunque anche $\bar{\theta}$: sono necessari degli stimatori per R .

Supponiamo che per ogni n sia definito uno stimatore di massima verosimiglianza $\hat{\theta}_n$, relativo al modello statistico

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\otimes_{i=1}^n f(x_i, \theta) | \theta \in \Theta\}),$$

ed inoltre che si abbia quasi certamente

$$\hat{\theta}_n(Y_1, \dots, Y_n) \rightarrow \bar{\theta}.$$

Questa assunzione può essere giustificata in questo modo. Per la legge forte dei grandi numeri si ha quasi certamente

$$\frac{1}{n} \sum_{i=1}^n \log(f(Y_i, \theta)) \rightarrow \int g(y) \log f(y, \theta) dy$$

Ora $\hat{\theta}(Y_1, \dots, Y_n)$ è un punto di massimo della variabile aleatoria

$$\frac{1}{n} \sum_{i=1}^n \log(f(Y_i, \theta)), \quad (7.5)$$

Delle condizioni sufficienti affinché la convergenza abbia luogo sono esplicitate in [33]. Nelle medesime condizioni in [33] si dimostra che lo stimatore di massima verosimiglianza può essere sviluppato come:

$$\hat{\theta}_n = \bar{\theta} + J^{-1} n^{-1} \sum_{i=1}^n u(Y_i, \bar{\theta}) + o_{\mathbb{P}}(n^{-\frac{1}{2}}), \quad (7.6)$$

dove

$$u(y, \theta) = \frac{\partial \log f(y, \theta)}{\partial \theta}$$

e la matrice J è definita come

$$J_{i,j} = \mathbb{E} \left\{ \frac{\partial^2 \log f(y, \theta)}{\partial \theta_i \partial \theta_j} \right\}.$$

Consideriamo ora la successione di variabili aleatorie

$$\int g(x) \log f(x, \hat{\theta}_n) dx \quad (7.7)$$

Supponendo che $\log f(x, \hat{\theta}_n)$ sia dominata, oppure che sia crescente, essa è uno stimatore fortemente consistente di $\int g(y) \log f(y, \bar{\theta}) dy$; non può tuttavia essere adottata perché dipende dalla funzione $g(x)$, che è incognita. Consideriamo il valore atteso Q_n

$$Q_n = \mathbb{E} \left\{ \int g(x) \log f(x, \hat{\theta}_n) dx \right\},$$

Il confronto fra diversi modelli statistici da adattare alla distribuzione di una variabile Y avverrà stimando Q_n optando alla fine per il modello che ne massimizza la stima. Resta sempre il problema di trovare uno stimatore per Q_n : definiamo allora

$$\hat{Q}_n = \frac{1}{n} \sum_{i=1}^n \log f(x, \hat{\theta}).$$

Akaike propone di adottare questo come stimatore, previo correggerne, almeno in parte, la distorsione nel modo seguente.

Abbiamo supposto che la successione $\hat{\theta}_n$ converga, supponendo che f sia di classe C^2 su \mathbb{R} , possiamo sviluppare in serie di Taylor attorno a $\bar{\theta}$ la funzione $\log(f(x, \hat{\theta}))$ fino al termine di second'ordine (ricordiamo che nei passaggi successivi non stiamo integrando rispetto alla variabile aleatoria $\hat{\theta}$), ottenendo:

$$R_n = \int g(y) \log(f(y, \hat{\theta})) dy = \int g(y) \left[\log f(y, \bar{\theta}) + \frac{\partial \log f(x, \bar{\theta})}{\partial \theta} (\hat{\theta} - \bar{\theta}) + \frac{1}{2} (\hat{\theta} - \bar{\theta})^t I(y, \bar{\theta}) (\hat{\theta} - \bar{\theta}) \right] dy + o_P(n^{-2})$$

dove

$$I(y, \theta) = \frac{\partial^2 \log f(y, \theta)}{\partial \theta_i \partial \theta_j}$$

Siccome la funzione (7.4) è differenziabile e $\bar{\theta}$ è un punto di minimo interno a Θ , esso deve essere stazionario. Supponendo che sia lecito commutare l'operazione di derivazione con quella di integrazione abbiamo perciò che

$$\int \frac{\partial \log f(x, \bar{\theta})}{\partial \theta} dx = 0;$$

dunque lo sviluppo, arrestato al secondo termine, si riduce a

$$R_n \sim Q_0 - \frac{1}{2n} (\hat{\theta} - \bar{\theta})^t J(y, \bar{\theta}) (\hat{\theta} - \bar{\theta}),$$

dove ricordiamo che

$$J(\theta)_{i,j} = \int g(y) I(y, \theta)_{i,j} dy.$$

In maniera analoga possiamo sviluppare la variabile aleatoria \hat{Q}_n arrendandoci al secondo termine dello sviluppo:

$$\begin{aligned}\hat{Q}_n &= \frac{1}{n} \sum_{i=1}^n \log f(Y_i, \hat{\theta}) = \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \log f(Y_i, \theta_0) + \frac{\partial \log f(Y_i, \bar{\theta})}{\partial \theta} + \frac{1}{2} (\hat{\theta} - \bar{\theta})^t I(Y_i, \bar{\theta}) (\hat{\theta} - \bar{\theta}) \right\}.\end{aligned}$$

Per la legge forte dei grandi numeri:

$$\frac{1}{n} \sum_{i=1}^n I(Y_i, \bar{\theta}) \xrightarrow{\mathbb{P}} J.$$

Quindi, arrendandoci al secondo termine, la differenza $\hat{Q}_n - R_n$ può essere sviluppata come

$$\hat{Q}_n - R_n = \frac{1}{n} \sum_{i=1}^n (\log f(Y_i, \theta_0) - Q_0) + \frac{1}{\sqrt{n}} \left(\frac{1}{n} \sum_{i=1}^n u(Y_i, \bar{\theta}) \right)^t \sqrt{n} (\hat{\theta} - \bar{\theta}) + o_{\mathbb{P}}(n^{-1}).$$

Per lo sviluppo (7.6) degli stimatori di massima verosimiglianza, per il teorema del limite centrale e quello di Slutsky si ha

$$\sqrt{n}(\hat{\theta} - \bar{\theta}) \xrightarrow{\mathcal{L}} J^{-1}Z,$$

dove $Z \sim N_p(0, K)$ e K è la matrice di covarianza del vettore aleatorio $u(Y, \theta_0)$. In virtù di questi risultati

$$(\sqrt{n}(\hat{\theta} - \bar{\theta}))^t J (\sqrt{n}(\hat{\theta} - \bar{\theta})) \xrightarrow{\mathcal{L}} Z^t J^{-1} Z. \quad (7.8)$$

Quindi (7.8) è una forma quadratica di un vettore gaussiano e, per il corollario 5.3.1, avrà media

$$\frac{\text{Tr} J^{-1} K}{n}.$$

Per il teorema 6.3.1 dunque

$$\hat{Q}_n - R_n \xrightarrow{\mathcal{L}} Z^t J^{-1} Z$$

e, approssimando il valore atteso con quello della distribuzione asintotica,

$$\mathbb{E}\{\hat{Q}_n - Q_n\} = \mathbb{E}\{\hat{Q}_n - R_n\} \sim \frac{\text{Tr}\{J^{-1}K\}}{n}.$$

A questo punto Akaiake adotta la parziale correzione

$$\hat{Q}_n - \frac{\text{Tr}J^{-1}K}{n}$$

come stimatore di Q_n . Resta da vedere cosa sia $\text{Tr}J^{-1}K$. Se $g(x) = f(x, \bar{\theta})$, se si impongono alcune condizioni ², è facile constatare che $J = K$ e dunque $\text{Tr}J^{-1}K = p$.

Moltiplicando per n si ottiene quindi il criterio di informazione di Akaiake.

7.3.5 Correzioni al criterio di informazione di Akaiake

In [18] si tenta di correggere ulteriormente il criterio di informazione di Akaiake. Poniamoci nel caso di un modello di regressione standard

$$Y = \mathbf{X}\beta + \sigma\epsilon$$

dove ϵ è un vettore di variabili indipendenti e identicamente distribuite aventi distribuzione gaussiana standard. In questo caso il criterio di informazione secondo Akaiake può essere riscritto come

$$AIC = -2n \log \hat{\sigma} - 2(p + 1) - n - n \log(2\pi),$$

dove $\hat{\sigma}$ indica lo stimatore di massima verosimiglianza di σ , mentre p indica la lunghezza del vettore dei parametri β (il fatto che sottraiamo $p + 1$ è dovuto al parametro aggiunto σ).

Nel paragrafo precedente non abbiamo visto la costruzione del criterio di

²Detto $\{\mathbb{P}^\theta | \theta \in \Theta\}$ il modello statistico definito dalla verosimiglianza $L(\theta, x)$, si deve supporre che la verosimiglianza sia di classe C^2 su \mathbb{R}^n ed inoltre che sia possibile invertire l'operatore di derivazione parziale con l'integrazione per funzioni del tipo

$$\theta \mapsto \mathbb{E}\{L(\theta)Y\}$$

dove Y è una funzione di quadrato integrabile.

informazione nel caso di regressione, essa è analoga a quella di un campione di variabili indipendenti ed identicamente distribuite: la variabile aleatoria

$$R_n = \frac{1}{n} \sum_{i=1}^n \int g(y|x_i) \log f(y|x_i, \hat{\beta}, \hat{\sigma}) dy$$

quantifica la differenza fra il modello di regressione proposto e quello reale; essa è stimata attraverso

$$\hat{Q}_n = \frac{1}{n} \sum_{i=1}^n \log f(Y_i|x_i, \hat{\beta}, \hat{\sigma})$$

che deve essere corretta, per diminuirne la distorsione.

D'ora in poi supporremo che il modello di regressione proposto coincida con quello reale. Per il teorema di Cochran, lo stimatore di massima verosimiglianza $\hat{\sigma}^2$ ha distribuzione

$$\hat{\sigma}^2 \sim \sigma^2 \chi_{n-p}^2$$

ed è indipendente da $\hat{\beta}$. In queste condizioni possiamo semplificare le espressioni di \hat{Q}_n e R_n :

$$\begin{aligned} \hat{Q}_n &= -\log \hat{\sigma} - \frac{1}{2} - \frac{1}{2} \log(2\pi) \\ R_n &= -\log \hat{\sigma} - \sum_{i=1}^n \frac{(\mathbf{X}_i^t \beta - \mathbf{X}_i^t \hat{\beta})^2 + \sigma^2}{\hat{\sigma}^2} - \frac{1}{2} \log(2\pi), \end{aligned}$$

dove \mathbf{X}_i indica l' i -esima riga della *design matrix*. In analogia con quanto esposto nel paragrafo precedente, per apportare la correzione desiderata analizziamo il valore atteso

$$\mathbb{E}\{\hat{Q}_n - R_n\} = -\frac{1}{2} + \frac{1}{2} \mathbb{E}\left\{ \frac{\sigma^2}{\hat{\sigma}^2} \left[\frac{1}{n\sigma^2} \sum_{i=1}^n (\mathbf{X}_i^t \hat{\beta} - \mathbf{X}_i^t \beta)^2 + 1 \right] \right\}.$$

In [18], con dei semplici passaggi, si arriva a determinare che

$$\mathbb{E}\{\hat{Q}_n - R_n\} = \frac{p+1}{n-p-2},$$

questo giustifica l'introduzione di una correzione al criterio di Akaiake, giungendo al cosiddetto criterio di Akaiake corretto, dato da

$$AIC_c = AIC + \frac{2(p+1)(p+2)}{n-p-2}$$

In questo paragrafo vogliamo esaminare un altro caso, che utilizzeremo in seguito: quello in cui nel modello di regressione standard

$$Y = \mathbf{X}\beta + \sigma\epsilon$$

la varianza σ^2 sia nota e dunque non necessiti di essere stimata a partire dai nostri dati. È necessario anche in questo caso correggere il criterio di informazione secondo Akaiake? Proseguiamo in maniera del tutto analoga a quanto fanno Claeskens e Hjort. Si ha

$$\hat{Q}_n = \frac{1}{n} \sum_{i=1}^n \log f(Y_i|x_i, \hat{\beta}, \sigma)$$

$$R_n = \frac{1}{n} \sum_{i=1}^n \int g(y|x_i) \log f(y|x_i, \hat{\beta}, \sigma) dy$$

Nel nostro caso perciò

$$\hat{Q}_n = -\log \sigma - \frac{1}{2n} \sum_{i=1}^n \frac{(Y_i - \mathbf{X}_i^t \hat{\beta})^2}{\sigma^2} - \frac{1}{2} \log(2\pi)$$

$$R_n = -\log \sigma - \frac{1}{2n} \sum_{i=1}^n \frac{(\mathbf{X}_i^t \beta - \mathbf{X}_i^t \hat{\beta})^2 + \sigma^2}{\sigma^2} - \frac{1}{2} \log(2\pi).$$

Consideriamone la differenza

$$\hat{Q}_n - R_n = \frac{1}{2n} \sum_{i=1}^n \frac{(\mathbf{X}_i^t \beta - \mathbf{X}_i^t \hat{\beta})^2 + \sigma^2 - (Y_i - \mathbf{X}_i^t \hat{\beta})^2}{\sigma^2} = \frac{1}{2n} \sum_{i=1}^n \frac{(\mathbf{X}_i^t \beta - \mathbf{X}_i^t \hat{\beta})^2}{\sigma^2} + \frac{1}{2} + \frac{\hat{\sigma}^2}{\sigma^2},$$

dove $\hat{\sigma}^2$ è lo stimatore di massima verosimiglianza di σ^2 in un modello di regressione standard con σ non noto. Vogliamo calcolare il valore atteso della variabile aleatoria $\hat{Q}_n - R_n$.

È dunque necessario studiare la differenza

$$\mathbf{X}\hat{\beta} - \mathbf{X}\beta;$$

si ha che

$$\mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t Y, \quad (7.9)$$

dal momento che si è supposto $Y = \mathbf{X}\beta + \epsilon$, possiamo riscrivere (7.9) come

$$\mathbf{X}\beta + \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \epsilon;$$

Indichiamo con H la matrice $\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$: essa è una matrice di proiezione, ossia $H \cdot H = I$. Quindi si ricava facilmente che

$$\frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i^t \hat{\beta} - \mathbf{X}_i^t \beta)^2 = \frac{1}{n} \|H\epsilon\|^2 = \frac{1}{n} \epsilon^t H \epsilon.$$

Grazie a quanto detto sulle forme quadratiche di vettori gaussiani sappiamo calcolare il valore atteso dell'ultima variabile aleatoria, pari a

$$\mathbb{E}\left\{\frac{1}{n} \epsilon^t H \epsilon\right\} = \frac{1}{n} \mathbb{E}\{H \epsilon \epsilon^t\} = \sigma^2 \frac{\text{Tr}H}{n} = \frac{p}{n} \sigma^2.$$

Per il teorema di Cochran sappiamo che gli stimatori $\hat{\beta}$ e $\hat{\sigma}^2$ sono indipendenti e che $\hat{\sigma}^2 \sim \sigma^2 \chi_{n-p}^2$, dunque

$$\mathbb{E}\{\hat{Q}_n - R_n\} = \frac{p}{n}.$$

Questo suggerisce che in questo caso non è necessario apportare ulteriori fattori correttivi al criterio di Akaiake.

7.3.6 Risultati ottenuti

Abbiamo visto che il criterio di Akaiake, sia nel caso di variabili identicamente distribuite, sia nel caso della regressione, è costruito supponendo l'indipendenza dell'osservazioni. Riprendiamo la terminologia utilizzata nel paragrafo 7.1: il processo $Y(t)$ relativo ad un paziente non può essere considerato a variabili indipendenti; al contrario la sua funzione di covarianza coincide con quella descritta nel capitolo precedente.

Si tratta dunque di creare, partendo dai nostri dati, un campionamento di variabili aleatorie indipendenti di regressione omoschedastiche, che abbiamo la stessa distribuzione delle variabili $Y(0), \dots, Y(4)$.

A questo si aggiunge il problema della differenza della taglia del campione per le diverse visite: restringendosi, per le ragioni discusse in 7.1, alle misurazioni antecedenti il 1987, abbiamo a disposizione un massimo di 5 osservazioni per individuo, classificate in base al tempo trascorso a partire dalla stima della sua data di sieroconversione. Tale stima avviene ponendo $t = 0$ per la prima visita in cui il paziente risulta sieropositivo: abbiamo dunque delle osservazioni delle variabili aleatorie $Y(0), \dots, Y(4)$. Nella tabella 7.3.6 riportiamo

	Anderson-Darling	Cramer-Von-Mises	Pearson	Kolmogorov-Smirnov
$t = 0$	0.074	0.065	0.161	0.107
$t = 1$	0.105	0.121	0.336	0.150
$t = 2$	0.325	0.392	0.538	0.150
$t = 3$	0.359	0.478	0.537	0.643
$t = 4$	0.421	0.378	0.727	0.397

Tabella 7.2: Stime ottenute con gli stimatori descritti.

la taglia del campione relativo a ciascun t .

t	0	1	2	3	4
Numero di osservazioni	199	177	139	66	13

Tabella 7.3: Taglia dei campioni relativi a diversi t .

Per ottenere un campionamento a variabili indipendenti con distribuzione uguale a quella delle variabili $Y(0), \dots, Y(4)$ da queste osservazioni, una volta fissate le 13 osservazioni dell'ultima visita, abbiamo estratto in maniera casuale 13 osservazioni da quelle della terza visita, da cui erano state preventivamente eliminate quelle relative agli individui della quarta visita. Procedendo fino alla visita relativa al tempo $t = 0$, abbiamo ottenuto per ciascun t , un campione di 13 osservazioni indipendenti. Ora prendiamo la media campionaria delle misurazione relative ad uno stesso tempo: in questo modo otteniamo delle osservazioni delle variabili aleatorie $Y'(0), \dots, Y'(4)$ indipendenti e gaussiane, con medie uguale rispettivamente alle medie di $Y(0), \dots, Y(4)$ e varianza pari a

$$\frac{\text{Var}Y(i)}{13}.$$

Vogliamo valutare la validità dell'ipotesi che la funzione delle medie sia lineare nel tempo, ossia dell'ipotesi secondo cui

$$\mu'(t) = \mu - \delta t.$$

Supponiamo che la funzione delle medie possa essere quanto meno ben approssimata da una funzione polinomiale, di grado minore o uguale a tre. Vogliamo allora stabilire quale sia il grado del polinomio con cui il modello di regressione si adatti meglio ai dati che abbiamo a disposizione: questo corrisponde di fatto a scegliere le variabili di regressioni, utilizzeremo dunque il criterio di informazione di Akaiake classico, perché considereremo la varianza delle variabili nota, dal momento che è già stata stimata nel capitolo precedente. Riportiamo dunque nella tabella 7.3.6 l'informazione secondo Akaiake e il valore massimizzato della log-verosimiglianza (indicato con L_n), rispetto al grado del polinomio interpolante.

Grado del polinomio	1	2	3
AIC	-22.9967	-16.569	-19.0915
L_n	-0.07184	4.334	4.0728

Tabella 7.4: Informazione secondo Akaiake (AIC) e valore massimizzato della log-verosimiglianza (L_n).

Dunque, confrontando i modelli con il criterio di informazione di Akaiake, sarebbe da preferire un modello quadratico. Notiamo però che la distinzione esistente fra il criterio di informazione circa il modello cubico e quello quadratico è essenzialmente dovuta alla penalizzazione del numero di parametri da stimare, perché i massimi delle log-verosimiglianze sono molto simili.

Resta comunque il fatto che i nostri dati contraddicono l'ipotesi di un trend lineare nel tempo che era stato proposto in [7]. Cercheremo di spiegare le possibili ragioni per cui questo accade nel paragrafo 7.3.7.

Una volta accettato un modello quadratico, procediamo ad una regressione per una stima dei coefficienti. Si tratta, nel nostro caso, di una regressione

standard. Detti a, b, c i coefficienti della funzione delle medie

$$\mu(t) = a + bt + ct^2$$

otteniamo, con lo stimatore dei minimi quadrati, le seguenti stime:

$$\hat{a} = 13.947 \quad \hat{b} = -0.601 \quad \hat{c} = 0.118$$

Le conclusioni a cui siamo giunti potrebbero essere state vizzate dal campionamento casuale effettuato per creare un campione a variabili indipendenti.

Proviamo perciò ad analizzare i soli dati riguardanti i pazienti che presentano 5 visite dalla sierconversione antecedenti il 1987. Abbiamo a disposizione 12 individui che soddisfano tali requisiti; facendo le medie delle osservazioni ad uguale tempo t_i , otteniamo anche in questo caso 5 osservazioni di variabili non più indipendenti; tuttavia se continuiamo ad ipotizzare che la struttura del second'ordine del processo coincida con quella dei sani, abbiamo già stimato la matrice di covarianza di queste variabili aleatorie.

In questa situazione accettiamo che il modello preferibile sia quadratico, e procediamo ad una stima dei parametri con lo stimatore di Aitken. Di seguito riportiamo le stime ottenute

$$\hat{a} = 13.1822 \quad \hat{b} = 0.2862 \quad \hat{c} = -0.0212$$

Se ora osserviamo il grafico 7.3.6 della funzione

$$f(t) = \hat{a} + \hat{b}t + \hat{c}t^2$$

possiamo notare che la funzione è addirittura crescente nell'intervallo $[0, 5]$: questo sembra essere in totale contrasto con il modello proposto in [7].

7.3.7 Ragioni del fallimento della convalida del modello

Riassumendo ciò che è stato determinato nel paragrafo precedente, l'analisi dei dati a nostra disposizione porta a concludere che per $t \in [0, 5]$ non solo la funzione delle medie $\mu(t)$ non sia lineare, ma che (cosa più allarmante) essa è crescente nell'intervallo considerato.

In questo paragrafo esponiamo delle motivazioni che potrebbero aver portato a tali conclusioni. Confrontiamo dunque il modello finora adottato, con

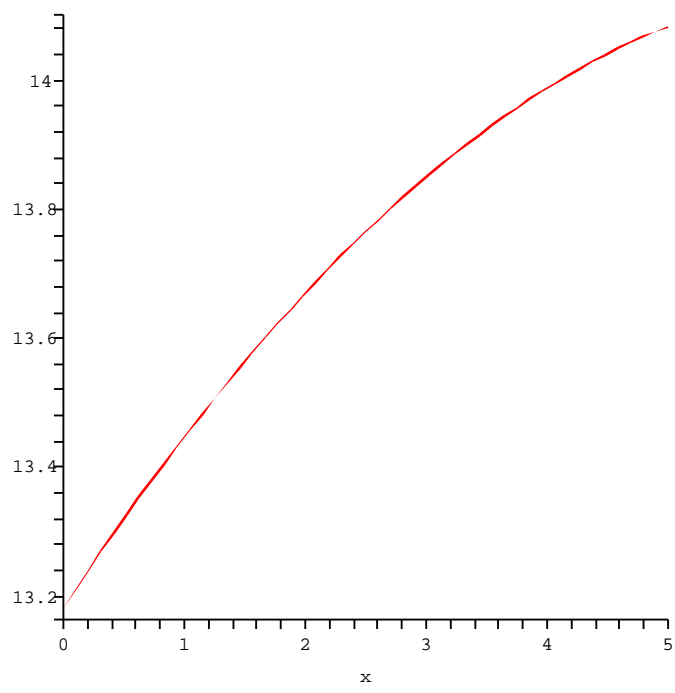


Figura 7.2: Grafico del polinomio $f(t)$.

quanto dice Carol Matthson Port in *Essentials of Pathophysiology: Concepts of Altered Health States* (Lippincott Williams and Wilkins, 2006) a proposito del decadimento di linfociti CD4+ dovuti all'infezione da HIV. Poiché non abbiamo competenze mediche, ci limitiamo a citare una parte del testo

Many persons, when they are initially infected with HIV, have an acute mononucleosis-like syndrome known as primary infection. This acute phase may include fever, fatigue, myalgias, sore throat, night sweats, gastrointestinal problems, lymphadenopathy, maculopapular rash, and headache [...]. During this time, there is an increase in viral replication, which leads to very high viral loads, sometime greater than 1,000,000 copies/mL, and a decrease in the CD4+ count. The signs and symptoms of primary HIV infection usually appear 2 to 4 weeks after exposure to HIV and last for a few days to 2 weeks. After several weeks, the immune system acts to control viral replication and reduces it to a lower level, where it remains for several years.

Questa è una grossa differenza con il modello proposto in [7]: esiste dunque, secondo Port, una fase iniziale, detta *infezione primaria*, in cui il numero di $cd4+$ per mm^3 di sangue diminuisce rapidamente per poi riportarsi ad un livello più alto. L'infezione primaria si svolgerebbe fra le due settimane e il mese dopo l'infezione da HIV e la ripresa sarebbe successiva, portando ad uno stadio più duraturo (vedi figura 7.3.7).

La nostra analisi sembra rilevare un andamento simile: questo spiegherebbe infatti sia perché la curva di regressione non sia lineare e, soprattutto, perché essa sia crescente.

D'altra parte però non possiamo né utilizzare la regressione determinata nel paragrafo precedente né prestare ulteriori attenzioni a questo periodo iniziale: il tempo trascorso fra una visita e la successiva (un semestre) è di molto maggiore rispetto a quello in cui avviene, ad esempio, l'infezione primaria; non è possibile dunque pensare di studiare la funzione delle medie in questo primo periodo approssimando con un range di sei mesi la data di sieroconversione.

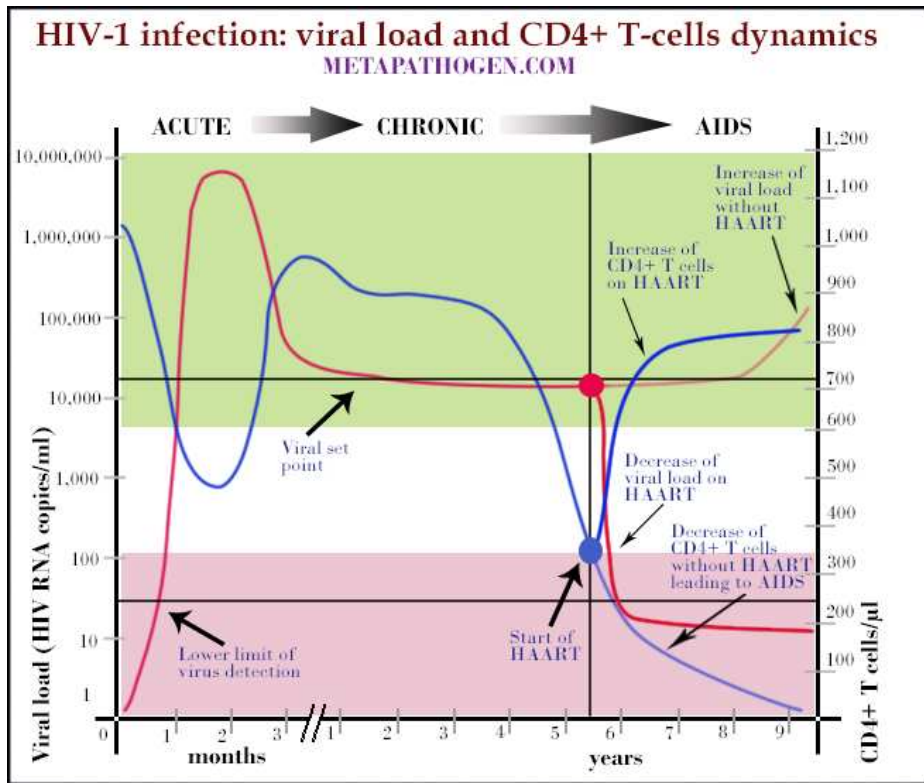


Figura 7.3: La linea blu indica il numero di linfociti CD4+, dal momento dell'infezione alla morte. Nel grafico risultano particolarmente evidenti l'infezione primaria (all'interno del primo anno di infezione) e la seguente ripresa dei linfociti cd4+, come spiegato da Port nel testo prima citato. L'immagine è tratta dal sito <http://www.metapathogen.com/HIV-1/HIV-1-disease-progression.html>

Capitolo 8

L'analisi dei dati dei sieropositivi

Nel capitolo precedente è emersa l'esistenza di un periodo che non era stato previsto in [7]: l'infezione primaria e la seguente fase di latenza del virus. In questo capitolo ci proponiamo di controllare se le ipotesi di [7] possano essere verificate nel periodo successivo a quello di latenza. Indichiamo, come nel capitolo precedente, il processo che modella il decadimento del numero di linfociti $CD4+$ nel sangue di una persona sieropositiva con

$$\{X(t)\}_{t \in \mathbb{R}^+},$$

vogliamo quindi verificare se nel periodo successivo alla latenza del virus la funzione delle medie $\mu(t)$ del processo gaussiano

$$\{f_\lambda(X(y))\}_{t \in \mathbb{R}^+} = \{Y(t)\}_{t \in \mathbb{R}^+}$$

è lineare nel tempo.

Il primo problema in questo tipo di analisi riguarda la natura dei dati a nostra disposizione: le nostre osservazioni sono realizzazioni del processo

$$\{Y(T + t)\}_{t \in \mathbb{R}^+}$$

dove T è la variabile aleatoria che rappresenta il tempo trascorso dall'infezione del virus alla prima visita del monitoraggio. Sulla variabile T non sappiamo di fatto nulla, perché non possiamo osservare alcuna sua realizzazione. Non abbiamo dunque alcuna indicazione diretta se i dati dei malati si

riferiscano al periodo successivo alla latenza o al periodo di latenza o, ancora, ad un intervallo temporale a cavallo fra i due periodi.

Per spiegare meglio il problema, riportiamo nel grafico seguente le misure relative a tre pazienti differenti, indentificati nell'archivio dai numeri 1254, 1797 e 9684. Per il secondo ed il terzo appare evidente che i trend sono, rispettivamente, crescente e decrescente: ma che cosa possiamo dire a proposito del primo? Supponiamo di poter attribuire le differenze fra i dati a

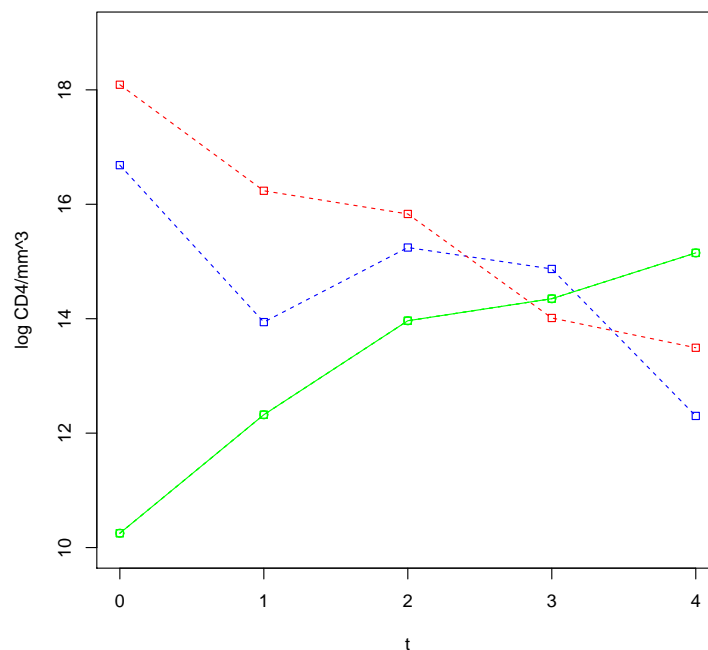


Figura 8.1: Il grafico riporta le misure in funzione del tempo del soggetto 1254(blu), 1797(verde) e 9684(rosso).

nostra disposizione solamente al periodo clinico che gli individui stanno attraversando: abbiamo la necessità, per proseguire nell'analisi, di isolare i dati che si riferiscono solo al periodo successivo a quello della latenza del virus. Gran parte della letteratura medica concorda sul fatto che, in questo periodo, la funzione media del processo $\{Y(t)\}_{t \in \mathbb{R}^+}$ è decrescente; eseguendo dunque una regressione non parametrica sulle osservazioni delle prime cinque visite di ogni paziente sieropositivo (le altre sono state scartate a causa della data

di commercializzazione del farmaco AZT) possiamo farci un'opinione, seppure approssimativa, riguardo alla decrescenza della funzione delle medie ¹.

Iniziamo con lo spiegare che cosa sia la regressione non parametrica, in base alla quale selezioniamo i soggetti che sembrano confermare un trend decrescente. Anche se questo modo di procedere non è completamente giustificabile, siamo costretti ad adottarlo per essere abbastanza sicuri di escludere i soggetti che stanno ancora attraversando il periodo di latenza del virus. In seguito, testiamo l'ipotesi della linearità del trend rispetto al tempo: introduciamo a questo scopo il test del T^2 di Hotelling. In base a questo test, i nostri dati sembrano rigettare l'ipotesi di linearità, anche per quanto riguarda i pazienti che hanno superato il periodo di latenza.

8.0.8 La regressione localmente polinomiale

In questo paragrafo vogliamo spiegare che cosa si intenda con regressione non parametrica e vedere un particolare metodo di procedere utilizzato in queste situazioni. Consideriamo due vettori aleatori: quello delle variabili dipendenti $Y = (Y_1, \dots, Y_n)$ e quello delle variabili indipendenti $X = (X_1, \dots, X_n)$. Come abbiamo fatto finora, considereremo il vettore delle variabili indipendenti come non aleatorio, e supporremo che sia verificato il modello di regressione

$$(Y_1, \dots, Y_n)^t = (f(X_1), \dots, f(X_n))^t + \epsilon^t,$$

dove il vettore degli errori $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ è un vettore aleatorio centrato e $f : \mathbb{R} \rightarrow \mathbb{R}$ (per semplicità considereremo solo il caso di variabili dipendenti e indipendenti a valori scalari). Nel caso del modello lineare generalizzato, descritto nel capitolo precedente, abbiamo ipotizzato che la funzione f fosse polinomiale; nel caso della regressione *non parametrica* non facciamo alcuna ipotesi globale su f .

Vediamo un possibile modo di procedere: in base a condizioni di regolarità che si suppongono soddisfatte, si opta per una regressione di natura locale; se si vuole dare una stima di $f(x_0)$, si approssima f in un intorno di x_0 mediante una particolare classe parametrica di funzioni.

Analizziamo il caso della regressione localmente polinomiale: essa trova le sue ragioni nel teorema di Taylor. Se supponiamo infatti che $f \in C^{k+1}(\mathbb{R})$,

¹Questo è il metodo più statistico di procedere: se avessimo maggiori competenze mediche potremmo analizzare sia il numero di linfociti $cd4+$, sia il livello della carica retrovirale o la comparsa delle prime malattie causate dall'infezione da HIV.

dato $x_0 \in \mathbb{R}$ è possibile avere un'approssimazione polinomiale di f in un determinato intorno di x_0

$$f(x) \approx \sum_{i=0}^k \frac{f^{(i)}(x_0)x^i}{i!} \quad \text{per } x \in (x_0 - \delta, x_0 + \delta).$$

La regressione localmente polinomiale mira ad una stima dei coefficienti $a_i(x_0) = \frac{f^{(i)}(x_0)}{i!}$, per poi giungere ad una stima di $f(x_0)$. Ma come avviene la stima dei coefficienti $a_i(x_0)$?

Supponiamo, per il resto del paragrafo, che il vettore degli errori ϵ sia composto da variabili indipendenti ed identicamente distribuite, con distribuzione gaussiana. Se avessimo ipotizzato che la funzione f fosse *globalmente* polinomiale, potremmo sfruttare il teorema di Gauss-Markov, che afferma che in queste condizioni uno stimatore ottimale nella classe degli stimatori corretti del vettore dei coefficienti (a_0, \dots, a_k) (essendo la funzione globalmente polinomiale, i coefficienti non hanno più natura locale) si ottiene minimizzando la funzione

$$E_{glob} = \sum_{i=1}^n (\hat{f}(X_i) - Y_i)^2.$$

Nel caso la funzione non sia globalmente polinomiale, ma sia ben approssimata in un intorno di x_0 da un polinomio di grado k in genere si procede in questo modo: per stimare il vettore $(a_0(x_0), \dots, a_k(x_0))$ si minimizza la funzione

$$E_{loc}(x_0) = \sum_{i=1}^n w\left(\frac{X_i - x_0}{h}\right) (\hat{f}_{x_0}(X_i) - Y_i)^2,$$

dove $w : [0, 1] \rightarrow \mathbb{R}$ è una particolare funzione che definisce il metodo di regressione localmente polinomiale, detta funzione *kernel*. Analogamente al caso del modello lineare generalizzato, utilizzando la notazione vettoriale definiamo la *design matrix* come

$$\mathbf{X} = \begin{pmatrix} 1 & (X_1 - x_0) & \dots & (X_1 - x_0)^k \\ 1 & (X_2 - x_0) & \dots & (X_2 - x_0)^k \\ \dots & \dots & \dots & \dots \\ 1 & (X_n - x_0) & \dots & (X_n - x_0)^k \end{pmatrix}$$

e la matrici dei coefficienti e dei pesi come

$$\mathbf{a} = \begin{pmatrix} a_0(x_0) \\ \dots \\ a_k(x_0) \end{pmatrix} \quad \mathbf{W} = \begin{pmatrix} w(X_1 - x_0) & 0 & \dots & 0 \\ 0 & w(X_2 - x_0) & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & w(X_n - x_0) \end{pmatrix}$$

La quantità da minimizzare si riduce a

$$(\mathbf{X}\mathbf{a} - Y)^t \mathbf{W} (\mathbf{X}\mathbf{a} - Y).$$

Differenziando si ricava un punto di minimo della funzione, dato da

$$\hat{\mathbf{a}} = (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} Y.$$

Dunque otteniamo una approssimazione di $f(x_0)$ come $\hat{\mathbf{a}}_0 = \hat{f}(x_0)$.

In accordo con quanto fa Cleveland in [49], tentiamo di isolare i fattori che caratterizzano un metodo di regressione locale, per sceglierne uno per il nostro caso. Riguardo al modo di procedere, abbiamo tre scelte importanti da fare: la scelta della funzione di *kernel*, quella del *fitting criterion* e della classe parametrica di funzioni approssimanti.

Per quanto riguarda la funzione di *kernel*, le scelte utilizzate sono le più svariate. Solitamente si opta per una funzione $w(u)$ che ha un massimo in $u = 0$ ed è decrescente per $|u|$ crescente. In realtà, ragioni di natura computazionale spingono a scegliere una funzione w che sia non nulla solo su un intervallo chiuso, in modo tale da considerare solo un numero finito di osservazioni. La funzione che definisce il metodo di *lowess* è la funzione tricubica, definita da

$$w(u) = \begin{cases} (1 - |u|^3)^3 & \text{se } |u| \leq 1 \\ 0 & \text{se } |u| > 1 \end{cases}$$

Un'alternativa è una funzione rettangolare, definita viceversa da

$$w(u) = \begin{cases} 1 & \text{se } |u| \leq 1 \\ 0 & \text{se } |u| > 1 \end{cases}$$

In [49] l'autore evidenzia il rapporto esistente fra la regolarità dell'approssimazione della funzione f e quella della funzione di *kernel*.

Il *fitting criterion* dipende invece dall'ipotesi fatta sulla distribuzione degli errori: nel caso esposto in precedenza, sembra giustificato utilizzare uno

stimatore simile a quello dei minimi quadrati; nel caso invece in cui la distribuzione degli errori ipotizzata sia diversa, è possibile scegliere altri *fitting criteria*, che potrebbero essere legati alla massimizzazione della verosimiglianza o ad altre tecniche statistiche.

Infine, c'è la scelta della classe parametrica di funzioni approssimanti: nel nostro caso abbiamo scelto la classe delle funzioni polinomiali (una scelta diversa potrebbe essere ad esempio quella dei polinomi trigonometrici); resta tuttavia da scegliere il grado dei polinomi e l'intervallo h .

Alcun suggerimenti sulla scelta di questi fattori, per la costruzione di un apposito metodo di regressione locale, possono essere trovati nello stesso [49], in [45] o in [22].

8.0.9 Metodo seguito nel nostro caso e risultati ottenuti

Nel nostro caso le cinque misurazioni delle visite relative ad uno stesso paziente sono realizzazioni delle variabili

$$(Y(T), \dots, Y(T + 4)) :$$

eseguiremo una regressione avente queste come variabili dipendenti e come variabile indipendente il tempo t . Detta $\mu(t)$ la funzione delle medie, possiamo riassumere la situazione nel modello di regressione non parametrica

$$(Y(T), \dots, Y(T + 4)) = (\mu(T), \dots, \mu(T + 4)) + \epsilon,$$

dove ϵ , vettore degli errori, ha struttura del second'ordine analoga a quella delle variabili $(Y(T), \dots, Y(T + 4))$. Dunque, nel nostro caso, le variabili che costituiscono il vettore dell'errore sono fra loro correlate ed abbiamo già stimato i valori della matrice di covarianza nel capitolo 5. Dal momento che abbiamo un numero esiguo di realizzazioni a disposizione (in tutto 5), scegliamo l'intervallo h del metodo in modo che esso sia centrato nel punto in cui si desidera la stima della funzione e che contenga i due punti simmetrici rispetto ad esso (siccome i valori della variabile indipendente sono equispaziati, questo equivale a porre $h = 1$). La ristrettezza del numero di osservazioni considerato per la stima locale della funzione ci costringe a scegliere un basso grado del polinomio: noi stimeremo i parametri sfruttando una approssimazione localmente lineare della funzione. Scegliamo infine la funzione *kernel*

rettangolare, cioè

$$w(u) = \begin{cases} 1 & \text{se } |u| \leq 1 \\ 0 & \text{se } |u| > 1 \end{cases}$$

A questo punto resta da scegliere il *fitting criterion*. Siccome lo stimatore dei coefficienti locali $(a_0(x_0), a_1(x_0))$ è lineare nelle osservazioni (y_1, \dots, y_n) , se supponiamo che la funzione μ sia ben approssimata negli intervalli $(x_0 - 1, x_0 + 1)$, sfruttiamo il teorema 7.3.1 per la stima locale dei coefficienti, utilizzando la stima della matrice di covarianza già data nel quarto capitolo. Il metodo può essere applicato solo a punti interni, cioè per le osservazioni delle variabili $Y(T + h), \dots, Y(T + 3h)$. Le approssimazioni della funzione μ nei punti $t = 0$ e $t = 4$ è stata ottenuta considerando, con metodo del tutto analogo, i due punti precedenti o seguenti.

Il grafico che segue riporta rappresenta la regressione svolta sul paziente 1254, di cui abbiamo già parlato.

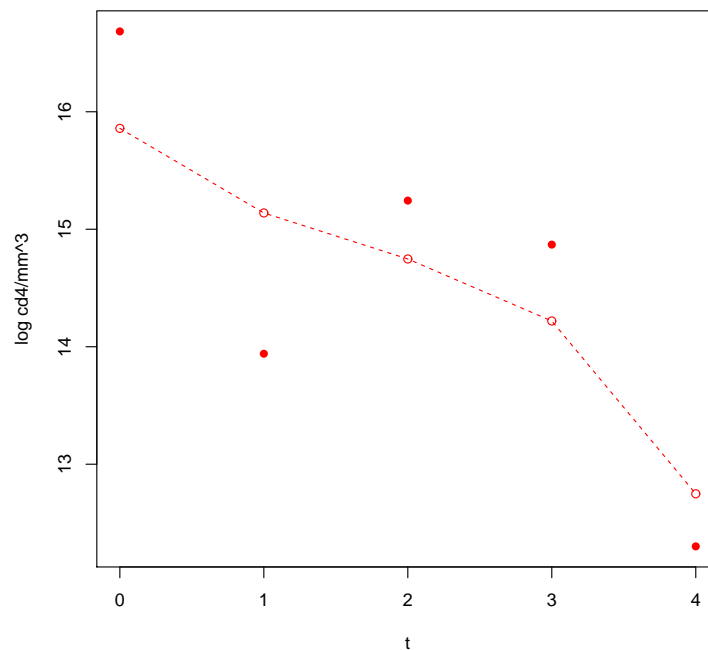


Figura 8.2: Il grafico riporta le misure relative al paziente (i punti pieni) e i valori ottenuti tramite la regressione (i punti vuoti)

Una volta svolta la regressione su ogni paziente, possiamo operare una selezione dei pazienti che adopereremo per proseguire l'analisi. Considereremo solo i dati di quei pazienti per cui la regressione locale genera una funzione decrescente: in questo modo riduciamo drasticamente il numero di individui utili all'analisi, che passa da 965^2 a 97. Ciò che ci interessa maggiormente, tuttavia, è avere una relativa certezza di aver selezionato pazienti appartenenti al giusto periodo, quindi accettiamo questa drastica riduzione della taglia del campione.

8.1 La verifica dell'ipotesi di linearità

Supponendo la bontà del metodo di estrazione del sottocampione seguito nei paragrafi precedenti, possiamo effettivamente tentare di testare se, nel periodo successivo a quello di latenza del virus, la media del processo $\{Y(t)\}_{t \in \mathbb{R}^+}$ decresce in modo lineare.

Supponiamo, per il momento, che ciò avvenga e seguiamo il metodo per la stima del coefficiente angolare della retta

$$\mu(t) = \mu_0 - \delta t$$

proposto in [7]. Consideriamo le variabili $Y(T + nh)$ e $Y(T + (n + 1)h)$ e il loro rapporto incrementale

$$R_n = \frac{Y(T + (n + 1)h) - Y(T + nh)}{h}.$$

La distribuzione del vettore aleatorio

$$(Y(T + (n + 1)h), Y(T + nh))$$

condizionale a $T = t$ è gaussiana, con vettore delle medie $(\mu_0 - n\delta t, \mu_0 - \delta t - \delta h)$ e matrice di covarianza

$$\begin{pmatrix} \Gamma(0) & \Gamma(h) \\ \Gamma(h) & \Gamma(0) \end{pmatrix}$$

²Il numero di pazienti sieropositivi considerati può apparire esiguo rispetto al numero di individui sani considerati nel capitolo 5, ma era già stato filtrato per considerare solamente i pazienti che si fossero presentati a tutte e 5 le visite prima del 1987.

dove abbiamo indicato con Γ la funzione di covarianza del processo $\{Y(n)\}_{t \in \mathbb{R}^+}$. Dunque la variabile R_n , condizionata a $T = t$, ha distribuzione gaussiana, con media $-\delta$ e varianza

$$\sigma_R^2 = \frac{2(\Gamma(0) - \Gamma(h))}{h^2}.$$

Poiché la legge di R_n condizionata a $T = t$ non dipende da t , R_n ha distribuzione (non condizionale) gaussiana, con media $-\delta$ e varianza σ_R^2 .

Anche i vettori aleatori $(Y(T), Y(T+h), \dots, Y(T+4h))$ hanno distribuzione condizionale a $T = t$ gaussiana, con media $(\mu, \nu - \delta ht, \dots, \nu - 4\delta ht)$ e matrice di covarianza

$$\begin{pmatrix} \Gamma(0) & \Gamma(h) & \dots & \Gamma(4h) \\ \dots & \dots & \dots & \dots \\ \Gamma(4h) & \Gamma(3h) & \dots & \Gamma(0) \end{pmatrix}$$

Quindi, in modo del tutto analogo a quanto esposto sopra, si verifica che, detto

$$R_i = \frac{Y(T + (i+1)h) - Y(T + ih)}{h},$$

i vettori

$$(R_0, \dots, R_3)$$

hanno distribuzione (non condizionale) gaussiana, con vettore delle medie

$$(-\delta, \dots, -\delta).$$

Questo ci dà contemporaneamente l'occasione di testare la linearità delle funzioni delle medie e di stimare, nel caso ciò risultasse possibile, il parametro di decadimento δ .

Per ognuno dei soggetti selezionati nella sezione precedente calcoliamo le differenze fra una visita e la precedente: otteniamo in questo modo, per ciascun individuo, un'osservazione del vettore aleatorio

$$(R_0, \dots, R_3).$$

Se il modello proposto fosse verificato, le osservazioni si (R_0, \dots, R_3) relative ad ogni paziente dovrebbero essere la realizzazione di un vettore gaussiano, con media $(-\delta, \dots, -\delta)$.

Abbiamo testato parzialmente l'ipotesi sulla distribuzione dei campioni: la cosa migliore che potremmo fare è testare l'ipotesi di gaussianità per l'intero

vettore (R_0, \dots, R_3) ; noi ci limiteremo tuttavia a testare separatamente la gaussianità delle variabili R_0, \dots, R_4 . I risultati sono riportati nella tabella 8.1.

	Anderson-Darling	Cramer-Von-Mises	Pearson	Kolmogorov-Smirnov
R_0	0.03966	0.04389	0.4333	0.01066
R_1	0.2628	0.2051	0.1355	0.07932
R_2	0.3510	0.385	0.3063	0.2800
R_3	0.3510	0.385	0.06233	0.252

Tabella 8.1: P-values relativi ai test di normalità eseguiti sulle osservazioni di R_i .

Con qualche riserva su R_0 , decidiamo di accettare l'ipotesi di gaussianità delle singole variabili R_i .

Supposto che anche la distribuzione congiunta di R_0, \dots, R_3 sia gaussiana, nel prossimo paragrafo vediamo come fare a testare l'ipotesi

$$\mathbb{E}\{R_0\} = \dots = \mathbb{E}\{R_3\}.$$

8.2 Il test sulla statistica di Hotelling

Per la costruzione del test basato sulla statistica di Hotelling abbiamo bisogno della definizione di matrice aleatoria.

Definizione 8.2.1. *Siano $X_{i,j}$ per $i = 1, \dots, n$ e $j = 1, \dots, m$ variabili aleatorie reali definite su $(\Omega, \mathcal{F}, \mathbb{P})$. La variabile aleatoria*

$$\Omega \rightarrow \mathcal{M}(m \times n, \mathbb{R})$$

$$\omega \mapsto \{A(\omega)\}_{i,j} = X_{i,j}(\omega)$$

dove $\mathcal{M}(m \times n, \mathbb{R})$ è l'insieme delle matrici di dimensione $m \times n$ a coefficienti reali, si dice una matrice aleatoria.

Identificando gli spazi $\mathcal{M}(m \times n, \mathbb{R})$ e $\mathbb{R}^{m \times n}$, possiamo pensare una matrice aleatoria come un vettore aleatorio: la probabilità immagine di una matrice aleatoria A sarà la probabilità immagine sullo spazio $(\mathbb{R}^{m \times n}, \mathcal{B}(\mathbb{R}^{m \times n}))$. Supponiamo che $\mathbf{X}_1, \dots, \mathbf{X}_n$ sia un campione di vettori aleatori definiti su $(\Omega, \mathcal{F}, \mathbb{P})$ indipendenti ed equidistribuiti, con

$$\mathbf{X}_i \sim N(\mu, Q).$$

Indichiamo con $\bar{\mathbf{X}}$ e \mathbf{S}_X , rispettivamente, le variabili aleatorie

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$$

e

$$\mathbf{S}_X = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^t = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t - n \bar{\mathbf{X}} \bar{\mathbf{X}}^t.$$

Queste variabili sono di fondamentale importanza per l'estensione al caso vettoriale del teorema di Cochran e per la derivazione del test di Hotelling.

8.2.1 La distribuzione di Wishart e la statistica di Hotelling.

Definizione 8.2.2. *Supponiamo che $\mathbf{X}_1, \dots, \mathbf{X}_n$ siano vettori aleatori, definiti su $(\Omega, \mathcal{F}, \mathbb{P})$, indipendenti ed equidistribuiti, con*

$$\mathbf{X}_i \sim N_m(0, Q).$$

Consideriamo la matrice aleatoria $A = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t$. La distribuzione di probabilità che la matrice A induce su $(\mathbb{R}^{m \times m}, \mathcal{B}(\mathbb{R}^{m \times m}))$ si dice distribuzione di Wishart centrata con parametri n e Q e si indica con $W_m(n, Q)$.

Dato un vettore $\mathbf{x} \in \mathbb{R}^n$, la matrice

$$\mathbf{x} \mathbf{x}^t$$

è semidefinita positiva, dunque la distribuzione di Wishart è concentrata sulle matrici semidefinite positive.

Proposizione 8.2.1. *Supponiamo che $\mathbf{X}_1, \dots, \mathbf{X}_n$ sia un campione di vettori aleatori indipendenti ed equidistribuiti, con legge*

$$\mathbf{X}_i \sim N_m(0, Q),$$

dove Q è definita positiva. La matrice aleatoria

$$\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t$$

è, almeno quasi certamente, invertibile.

Dimostrazione. Lo dimostriamo solo nel caso $n = 1$: nel caso $n > 1$ la dimostrazione è del tutto analoga. Supponiamo dapprima che $Q = I_m$. In questo caso, le componenti del vettore aleatorio $\mathbf{X} := \mathbf{X}_1$ sono indipendenti con distribuzione normale standard. Imponiamo che la matrice aleatoria $\mathbf{X}\mathbf{X}^t$ abbia rango minore di m : dunque per ogni $j = 1, \dots, m$ esiste f_j tale che, indicando con \mathbf{X}^j la j -esima componente del vettore aleatorio \mathbf{X}

$$\mathbf{X}^j = f_j(\mathbf{X}^1, \dots, \mathbf{X}^{j-1}, \mathbf{X}^{j+1}, \dots, \mathbf{X}^m).$$

Questo evento ha probabilità nulla, visto che le componenti di \mathbf{X} sono indipendenti. Ora supponiamo che Q sia semplicemente definita positiva: esistono una matrice definita positiva B e un vettore aleatorio gaussiano \mathbf{Y} , con $\mathbf{Y} \sim N_m(0, I)$ tali che

$$\mathbf{X} = B\mathbf{Y}.$$

Dunque

$$\mathbf{X}\mathbf{X}^t = B\mathbf{Y}\mathbf{Y}^t B^t;$$

dal momento che le matrici B e B^t sono non singolari e $\mathbf{Y}\mathbf{Y}^t$ non lo è, almeno quasi certamente, per quanto detto sopra, concludiamo che anche la matrice $\mathbf{X}\mathbf{X}^t$ è, quasi certamente, non singolare. \square

Dunque se Q è una matrice quadrata di dimensione n definita positiva e simmetrica, la distribuzione di Wishart con parametro Q e $n \in \mathbb{N}$ è concentrata sulle matrici definite positive. Questo garantisce la correttezza della definizione che segue.

Definizione 8.2.3. Supponiamo che \mathbf{X} sia un vettore aleatorio definito su $(\Omega, \mathcal{F}, \mathbb{P})$ e a valori in \mathbb{R}^m e che A sia una matrice aleatoria definita su $(\Omega, \mathcal{F}, \mathbb{P})$ con

$$A \sim W_m(n, Q) \quad \mathbf{X} \sim N_m(\mu, Q).$$

Se A e \mathbf{X} sono indipendenti e Q è non singolare, la distribuzione della variabile aleatoria

$$T^2 = \mathbf{X}^t A^{-1} \mathbf{X}$$

si dice distribuzione T^2 con parametro di non centralità:

$$\delta = \mu^t Q^{-1} \mu,$$

e gradi di libertà n e m . Essa si indica con il simbolo $T_{m,n,\delta}^2$.

Si può dimostrare (vedi ad esempio [40]), che, se $Y \sim T_{m,n,\delta}^2$, la variabile aleatoria

$$\frac{n-m+1}{nm} Y \sim F_{m,n-m+1,\delta},$$

dove F indica la distribuzione di Fisher-Snedecor.

8.2.2 La versione vettoriale del teorema di Cochran

Per dimostrare la versione vettoriale del teorema di Cochran sono necessari due lemmi di algebra lineare e una proposizione.

Lemma 8.2.1. Supponiamo che $\{c_{i,j}\}_{i,j=1,\dots,n}$ sia una matrice ortogonale a coefficienti reali e \mathbf{x}_i per $i = 1, \dots, n$ siano vettori in \mathbb{R}^m . Consideriamo, per $i = 1, \dots, n$, i vettori \mathbf{y}_i , definiti da

$$\mathbf{y}_i = \sum_{j=1}^n c_{i,j} \mathbf{x}_j,$$

allora vale

$$\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t = \sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^t.$$

Dimostrazione. Si tratta semplicemente di un calcolo:

$$\sum_{i=1}^n \mathbf{y}_i \mathbf{y}_i^t = \sum_{i=1}^n \left(\sum_{s=1}^n c_{i,s} \mathbf{x}_s \right) \left(\sum_{t=1}^n c_{i,t} \mathbf{x}_t \right)^t =$$

$$\begin{aligned}
&= \sum_{t,s=1}^n \left(\sum_{i=1}^n c_{i,t} c_{i,s} \right) \mathbf{x}_t \mathbf{x}_s^t = \sum_{t,s=1}^n \delta_{s,t} \mathbf{x}_t \mathbf{x}_s^t = \\
&= \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t.
\end{aligned}$$

dove $\delta_{s,t}$ è la funzione *delta* di Kronecker. □

Lemma 8.2.2. *Sia A una matrice a coefficienti reali di dimensioni $n \times m$, di rango m , con $m < n$, tale che*

$$A^t A = I_m,$$

dove I_m è la matrice identità di dimensione m . Esiste una matrice B di dimensioni $n \times (m - n)$, tale che la matrice quadrata

$$[AB]$$

è una matrice ortogonale.

Dimostrazione. Poiché A ha rango m , esiste una matrice C di dimensioni $n \times (n - m)$ tale che la matrice $[AC]$ ha rango massimo. Consideriamo la matrice D definita da

$$D = C - AA^t C.$$

Essa verifica $D^t A = 0$, infatti

$$D^t A = (C^t - C^t AA^t) A = C^t A - C^t A = 0,$$

inoltre $D^t D$ è una matrice definita positiva, perché ha rango massimo, dunque esiste una matrice ortogonale E tale che

$$ED^t DE = I.$$

Consideriamo allora la matrice $B = DE$: è facile vedere che $[AB]$ è ortogonale. □

Proposizione 8.2.2. *Supponiamo che $\mathbf{X}_1, \dots, \mathbf{X}_n$ siano vettori gaussiani aleatori e indipendenti, dove $\mathbf{X}_i \sim N_m(\mu_i, Q)$; se $\{c_{i,j}\}_{i,j=1,\dots,n}$ è una matrice ortogonale, i vettori aleatori*

$$\mathbf{Y}_i = \sum_{j=1}^n c_{i,j} \mathbf{X}_j$$

sono ancora indipendenti e gaussiani, con

$$\mathbf{Y}_i \sim N_m(\nu_i, Q),$$

dove $\nu_i = \sum_{j=1}^n c_{i,j} \mu_j$.

Dimostrazione. Indichiamo con $[\mathbf{X}_i]_j$ la j -esima componente del vettore aleatorio \mathbf{X}_i ; allora l'insieme di variabili aleatorie

$$\{[\mathbf{X}_i]_j\}_{i,j=1,\dots,n}$$

è un sistema gaussiano: di conseguenza anche i vettori \mathbf{Y}_i sono vettori gaussiani. Calcoliamone il vettore delle medie:

$$\mathbb{E}\{\mathbf{Y}_i\} = \mathbb{E}\left\{\sum_{j=1}^n c_{i,j} \mathbf{X}_i\right\} = \sum_{j=1}^n c_{i,j} \mu_j = \nu_i.$$

Poiché l'insieme $\{[\mathbf{X}_i]_j\}_{i,j=1,\dots,n}$ è un sistema gaussiano, per dimostrare l'indipendenza fra i vettori \mathbf{Y}_i , ci basta dimostrare che se $i_1 \neq i_2$ per ogni $j_1, j_2 = 1, \dots, n$ si ha

$$\text{cov}\{[\mathbf{X}_{i_1}]_{j_1}, [\mathbf{X}_{i_2}]_{j_2}\} = 0.$$

Dal momento che la matrice $\{c_{i,j}\}_{i,j=1,\dots,n}$ è ortogonale,

$$\begin{aligned} \text{cov}\{[\mathbf{X}_{i_1}]_{j_1}, [\mathbf{X}_{i_2}]_{j_2}\} &= \text{cov}\left\{\left(\sum_{t=1}^n c_{i_1,t} [\mathbf{X}_t]_{j_1}\right) \left(\sum_{s=1}^n c_{i_2,s} [\mathbf{X}_s]_{j_2}\right)\right\} = \\ &= \sum_{s,t=1}^n c_{i_1,t} c_{i_2,s} \mathbb{E}\{([\mathbf{X}_t]_{j_1} - [\mu_t]_{j_1})([\mathbf{X}_s]_{j_2} - [\mu_s]_{j_2})\} = \\ &= \sum_{s,t=1}^n \delta_{t,s} c_{i_1,t} c_{i_2,s} q_{j_1,j_2} = q_{j_1,j_2} \sum_{t=1}^n c_{i_1,t} c_{i_2,t} = \delta_{i_1,i_2} q_{j_1,j_2}, \end{aligned}$$

dove $Q = \{q_{i,j}\}_{i,j=1,\dots,m}$. Questo dimostra che, se $i_1 \neq i_2$, i vettori aleatori \mathbf{Y}_{i_1} e \mathbf{Y}_{i_2} sono indipendenti, mentre se $i_1 = i_2$, la matrice di covarianza di \mathbf{Y}_{i_1} è Q . \square

Proviamo la versione vettoriale del teorema di Cochran.

Proposizione 8.2.3. Sia $\mathbf{X}_1, \dots, \mathbf{X}_n$ un campione di vettori aleatori indipendenti e gaussiani, con

$$\mathbf{X}_i \sim N_m(\nu, Q)$$

allora le variabili aleatorie $\bar{\mathbf{X}}$ e \mathbf{S}_X , definite dopo la definizione 8.2.1, sono indipendenti. Inoltre

$$\bar{\mathbf{X}} \sim N(\nu, \frac{1}{n}Q)$$

$$\mathbf{S}_X \sim W_m(n-1, Q)$$

Dimostrazione. Consideriamo il vettore

$$\mathbf{a} = \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)^t,$$

dal momento che $\mathbf{a}^t \mathbf{a} = 1$, per il lemma 8.2.1 esiste una matrice ortogonale B , di dimensione n , che ha come ultima riga proprio il vettore \mathbf{a}^t .

Definiamo i vettori aleatori, per $i = 1, \dots, n$

$$\mathbf{Z}_i = \sum_{j=1}^n b_{i,j} \mathbf{X}_j.$$

Per la proposizione 8.2.2 i vettori \mathbf{Z}_i sono ancora gaussiani, indipendenti ed equidistribuiti con matrice di covarianza Q e media $\sum_{j=1}^n b_{i,j} \mu_j$.

La variabile aleatoria $\mathbf{Z}_n = \sum_{j=1}^n b_{n,j} \mathbf{X}_j$ coincide con la variabile $\sqrt{n} \bar{\mathbf{X}}$. Inoltre, poiché B è una matrice ortogonale, l'applicazione del lemma 8.2.1 garantisce che

$$\begin{aligned} \mathbf{S}_X &= \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^t = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^t - n \bar{\mathbf{X}} \bar{\mathbf{X}}^t = \\ &= \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^t - \mathbf{Z}_n \mathbf{Z}_n^t = \sum_{i=1}^{n-1} \mathbf{Z}_i \mathbf{Z}_i^t. \end{aligned}$$

Dal momento che i vettori \mathbf{Z}_i sono indipendenti, anche le variabili aleatorie \mathbf{S}_X e $\bar{\mathbf{X}}$ sono indipendenti. Visto che $b_{n,i} \sqrt{n} = 1$, si ha

$$\mathbb{E}\{\mathbf{Z}_i\} = \sum_{j=1}^n b_{i,j} \mathbb{E}\{\mathbf{Z}_j\} = \sum_{j=1}^n b_{i,j} \mu = \sum_{j=1}^n b_{i,j} b_{n,j} \sqrt{n} \mu = 0$$

se $i \neq n$, quindi per definizione di distribuzione di Wishart

$$\mathbf{S}_X \sim W_m(n-1, Q).$$

□

8.2.3 Test di ipotesi sulla media di un vettore gaussiano

In questo paragrafo costruiamo un test sulla media di un vettore gaussiano. Supponiamo che $\mathbf{X}_1, \dots, \mathbf{X}_n$ sia un campione di vettori gaussiani ed equidistribuiti, con media μ e matrice di covarianza Q , a valori in \mathbb{R}^m . Data una matrice C di dimensioni $c \times m$ di rango c , con $c < m$, vogliamo testare l'ipotesi nulla H_0 contro l'alternativa H_1 così specificate:

$$H_0 : C\mu = 0 \quad H_1 : C\mu \neq 0.$$

Consideriamo i vettori aleatori \mathbf{Z}_i a valori in \mathbb{R}^c , definiti da

$$\mathbf{Z}_i = C\mathbf{X}_i.$$

Essi sono ancora vettori indipendenti e gaussiani, con

$$\mathbf{Z}_i \sim N_c(C\mu, CQC^t).$$

Per la proposizione 8.2.3

$$\sqrt{n}\mathbf{Z} \sim N_c(\sqrt{n}C\mu, CQC^t)$$

e

$$\begin{aligned} \mathbf{S}_z &= \frac{1}{n} \sum_{i=1}^n (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^t = \\ &= \sum_{i=1}^n (C\mathbf{X}_i - C\bar{\mathbf{X}})(C\mathbf{X}_i - C\bar{\mathbf{X}})^t = \\ &= \sum_{i=1}^n C(\mathbf{X}_i - \bar{\mathbf{X}})(C(\mathbf{X}_i - \bar{\mathbf{X}}))^t = \sum_{i=1}^n C(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^t C^t = C\mathbf{S}_X C^t \sim W_c(n-1, CQC^t). \end{aligned}$$

La stessa proposizione 8.2.3 assicura che le variabili aleatorie \mathbf{S}_Z e $\bar{\mathbf{Z}}_n$ sono indipendenti. Dal momento che la matrice C ha, per ipotesi, rango c , la matrice

$$CQC^t$$

è una matrice invertibile: dunque la variabile aleatoria

$$T^2 = (\sqrt{n}\bar{\mathbf{Z}})^t \mathbf{S}_z^{-1} (\sqrt{n}\bar{\mathbf{Z}}) = n(C\bar{\mathbf{X}})^t (CSC^t)^{-1} (C\bar{\mathbf{X}})$$

ha distribuzione $T_{c,n-1,\delta}^2$, dove il parametro di non centralità δ è dato da

$$\delta = n(C\mu)^t (CQC^t)^{-1} (C\mu).$$

La variabile aleatoria

$$F = \frac{(n-1) - c + 1}{(n-1)c} T^2 = \frac{n-c}{(n-1)c} T^2$$

ha distribuzione di Fisher-Snedecor con parametri c , $n-c$ e δ . Se vale l'ipotesi nulla, ossia se $C\mu = 0$, si ha

$$F \sim F_{c,n-c},$$

dunque il test con regione di rifiuto

$$\{F > \Psi(1 - \alpha)\}$$

(dove Ψ indica l'inversa generalizzata della funzione di ripartizione di $F_{c,n-c}$) ha taglia α .

Vediamo come si applica questo test al nostro caso: abbiamo un campione di vettori gaussiani indipendenti

$$\mathbf{X}_i \sim N_m(\mu, Q)$$

con $\mu = (\mu_1, \dots, \mu_m)$. Vogliamo testare l'ipotesi nulla

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_m.$$

Consideriamo la matrice C_0 di dimensione $m-1 \times m$ definita da

$$C_0 = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix}$$

L'ipotesi nulla H_0 è equivalente all'ipotesi

$$C_0\mu = 0,$$

quindi si può procedere nel modo indicato precedentemente.

8.3 Risultati ottenuti e possibili modifiche al modello

Calcolando le differenze fra le misure di visite successive in uno stesso paziente, otteniamo un campionamento del vettore, che supponiamo gaussiano,

$$(R_0, \dots, R_3) :$$

con il test basato sulla statistica T^2 di Hotelling possiamo testare l'ipotesi nulla

$$H_0 : \mathbb{E}\{R_0\} = \dots = \mathbb{E}\{R_3\}.$$

Il test rigetta l'ipotesi nulla, con un *p-value* di

$$p = 0.0000329.$$

Non possiamo quindi accettare l'ipotesi che la funzione delle medie del processo $\{Y(t)\}_{t \in \mathbb{R}^+}$ decresca linearmente nel tempo; perciò risulta infondato qualsiasi tentativo di stima del parametro di decadimento δ .

Gli autori di [7] prendono in considerazione questa ipotesi e propongono una parametrizzazione più generale per la funzione delle medie data da

$$\mu(t) = \mu \exp\left\{-\left(\frac{t}{\theta}\right)^\alpha\right\},$$

dove μ è la media nella popolazione sana, già stimata nel capitolo 4, e $\theta \in \mathbb{R}^+$, $\alpha \in \mathbb{R}^+$ sono parametri da stimare.

Sempre in [7], si propone l'utilizzo degli stimatori di massima verosimiglianza, all'interno dei dati di un singolo soggetto: noi abbiamo già dovuto scartare molte delle misurazioni disponibili, riducendoci a sole cinque misurazioni per soggetto. Non avrebbe più alcun senso utilizzare degli stimatori di massima verosimiglianza su un campione così ridotto; per questo ci fermiamo.

Capitolo 9

Conclusioni

I risultati originali di questa tesi sono, da una parte, la dimostrazione della forte consistenza della successione di stimatori proposta nel capitolo 4, e dall'altra, alcune modifiche proposte al modello di [7].

Il problema affrontato nel capitolo 4 è il seguente: data la variabile X , che modella il numero di linfociti $cd4+$ per mm^3 di sangue in un paziente sano, e supposto che esista $\alpha \in (0, 1)$ tale che X^α abbia legge gaussiana, ci proponiamo di stimare α .

Il metodo più diffuso in letteratura, quello di Cox-Box, si basa su ipotesi che non possono essere verificate (la distribuzione delle variabili considerate non può essere gaussiana) e questo rende molto difficile derivare dei risultati asintotici per gli stimatori proposti.

Il metodo che noi abbiamo proposto nella seconda parte del quarto capitolo non risolve esattamente lo stesso problema: noi abbiamo supposto che, data una particolare distribuzione di probabilità \mathbb{P}_0 , esista α tale che X^α abbia legge \mathbb{P}_0 e abbiamo proposto una successione di stimatori $\hat{\alpha}_n$ per α . Sotto alcune ipotesi su X e su \mathbb{P}_0 , abbiamo dimostrato che $\{\hat{\alpha}_n\}_{n \in \mathbb{N}}$ è una successione fortemente consistente di α .

L'ipotesi sulla variabile X è equivalente a supporre che il processo empirico di variabili indipendenti ed equidistribuite, aventi legge uguale a quella di X , converga puntualmente. Non abbiamo approfondito che cosa comporti quest'ipotesi, di cui si potrebbe probabilmente fare a meno, se ci si accontentasse della dimostrazione della consistenza debole per la successione di stimatori $\{\hat{\alpha}_n\}_{n \in \mathbb{N}}$.

Il nostro metodo risolve il problema nel caso in cui l'obiettivo della trasfor-

mazione è una specifica distribuzione di probabilità e *non* una famiglia di distribuzioni: resta un problema aperto dimostrare gli stessi risultati per la successione $\hat{\alpha}_n$ nel caso in cui si voglia che la legge di X^α appartenga ad una famiglia parametrica di leggi di probabilità. In questo caso, bisogna considerare il test di Kolmogorov-Smirnov con la correzione apportata da Lilliefors, perciò procedere è più complesso.

Per quanto riguarda i punti in cui sembra che il modello di [7] debba essere modificato, precisiamo che non siamo riusciti a reperire i dati su cui si erano basati gli autori per la formulazione del modello: non possiamo sapere, quindi, se le modifiche da apportare sono dovute a fenomeni effettivamente trascurati in [7] o a differenze fra il campione esaminato in [7] e i nostri dati. I punti in cui sembra che i nostri dati non siano in accordo col modello proposto sono sostanzialmente tre.

Per prima cosa, abbiamo proposto una nuova trasformazione T , alternativa al logaritmo, per cui si può assumere che la variabile trasformata $T(X)$ sia gaussiana. Siamo consapevoli che questa trasformazione non ha valore universale: lo stesso metodo seguito nella tesi conduce a trasformazioni radicalmente diverse se si adoperano campioni differenti.

Inoltre, sembra che gli autori di [7] abbiano effettivamente trascurato il fenomeno del *periodo di latenza*, con la conseguente *fase di latenza* del virus. La letteratura medica che abbiamo consultato concorda sull'esistenza di queste fasi nell'infezione da HIV: ciò porta ad escludere in partenza l'ipotesi di [7], secondo cui la media del livello (trasformato) di $cd4+$ decrescerebbe linearmente dal momento dell'infezione.

Infine, basandoci sui nostri dati, sembra che si possa *scartare* anche l'ipotesi che la media del livello (trasformato) di $cd4+$ decresca linearmente *dopo* la fase di latenza. Tuttavia, per arrivare a delle conclusioni più certe su questo punto occorrerebbero una specifica ricerca medica e ulteriori dati sui pazienti.

In particolare, è da segnalare a questo proposito che nel nostro database non sono presenti indicazioni riguardo all'assunzione dei farmaci nel periodo studiato: per questo abbiamo dovuto escludere ogni misura ottenuta in un periodo posteriore alla commercializzazione dei primi farmaci contro l'AIDS. Inoltre, abbiamo selezionato i soggetti che stanno attraversando il periodo posteriore a quello della latenza del virus in base ad una regressione locale sulle misure relative ad uno stesso individuo: non sappiamo quanto ciò sia giustificato; la taglia del campione considerato ne viene molto ridotta.

Ringraziamenti

Ringrazio per la disponibilità e l'aiuto che mi hanno dato Christopher Cox, Janet Schollemberg, Lisa Jacobson, Maria Cristina Prati e Enrico Peruzzo.

Ringrazio la mia famiglia ed in particolare Andrea, che ha migliorato i miei rapporti turbolenti con \LaTeX .

Ringrazio Ivan, che così sarà costretto ad accompagnarmi in Svezia.

Ringrazio Paola, perché mi ha ringraziato nella sua tesi.

Ringrazio Nicoletta, che ha rallegrato i miei week-end trentini.

Bibliografia

- [1] J. Wellner A. van der Vaart. *Weak Convergence and Empirical Processes*. Springer, 1996.
- [2] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- [3] T. Amemiya. Generalized least squares with an estimated autocovariance function. *Econometrica*, 41(4):723–732, July 1973.
- [4] T. Amemiya. *Advanced Econometrics*. Harvard University Press, 1985.
- [5] P. Baldi. *Calcolo delle probabilità e statistica*. Mc Graw Hill, 1998.
- [6] C.K. I. Williams C. E. Rasmussen. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [7] C. K. Sleeman C. J. Mode. *Stochastic processes in epidemiology: HIV/AIDS, other infectious diseases, and computers*. World Scientific, 2000.
- [8] R. G. E. Pinch C. M. Goldie. *Communication theory*. Cambridge University Press, 1991.
- [9] G. Matfin C. M. Porth. *Essentials of Pathophysiology: Concepts of Altered Health States*. Lippincott Williams and Wilkins, 2010.
- [10] H. Toutenburg C. R. Rao. *Linear models and generalizations: least squares and alternatives*. Springer, 2008.
- [11] A. Cohen. On estimating the mean and the standard deviation of truncated normal distributions. *Journal of the American Statistical Association*, 45:518–525, 1949.

- [12] H. Cramer. *Mathematical methods of statistics*. Princeton University Press, 1971.
- [13] J. W. Tuckey D. C. Hoaglin, F. Mosteller. *Fundamentals of Exploratory Analysis of Variance*. John Wiley and Sons, 1991.
- [14] J.L. Doob. *Stochastic Processes*. Wiley, 1990.
- [15] N. R. Draper and D. R. Cox. On distributions and their transformation to normality. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31:472–476, 1969.
- [16] J. P. Romano E. L. Lehmann. *Testing statistical hypotheses*. Springer, 2005.
- [17] W. Feller. *An introduction to probability theory and its applications*. Wiley and sons, 1971.
- [18] N. L. Hjort G. Claeskens. *Model selection and model averaging*. Cambridge University Press, 2008.
- [19] D. R. Cox G. E. P. Box. An analysis of transformation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26:211–252, 1964.
- [20] A. C. Harvey. *The econometric analysis of time series*. MIT Press, 1990.
- [21] J. T. Hattaway. *Parameter Estimation and Hypothesis Testing for the Truncated Normal Distribution with Applications to Introductory Statistics Grades*. 2010.
- [22] W. Härdle. *Applied nonparametric regression*. Cambridge University Press, 1992.
- [23] O. A. Georgewill; J. Ikimalo. Effect of azidothymidine on cd4 positive t cells in hiv positive patients. *Journal of Applied Sciences and Environmental Management*, 8(2):35–37, 2004.
- [24] J. P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3):419–426, December 1961.
- [25] R. Modarres J. Freeman. Inverse box cox: The power-normal distribution. *Statistics and Probability Letters*, 76:764–772, 2006.

- [26] P. Protter J. Jacod. *Probability Essentials*. Springer, 2004.
- [27] S.J. Taylor J.F.C. Kingman. *Introduction to measure and probability*. Cambridge University Press, 1977.
- [28] D. R. Anderson K. P. Burnham. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, 2002.
- [29] H. Kuo. *Introduction to stochastic integration*. Birkhaeuser, 2006.
- [30] S. L. Zeger L. S. Magder. A smooth nonparametric estimate of a mixing distribution using mixtures of gaussians. *Journal of the American Statistical Association*, 91:1141–1151, 1996.
- [31] C.A. Gotway L.A. Waller. *Applied Spatial Statics for Public Health Data*. Wiley Ieee Press, 2004.
- [32] M. Loeve. *Probability Theory, Volume 2*. Springer, 1978.
- [33] D. Pollard N. L. Hjort. Asymptotics for minimisers of convex processes. Technical report, University of Oslo, Oslo, Finland, 1993.
- [34] J. N. S. Matthews P. Armitage, G. Berry. *Statistical methods in medical research*. Wiley-Blackwell, 2002.
- [35] R. A. Davis P. J. Brockwell. *Time Series: Theory and Methods*. Springer, 1991.
- [36] An. V. Metcalfe P. S. P. Cowpertwait, A. Metcalfe. *Introductory Time Series with R*. Springer, 2009.
- [37] D. Pollard. *Convergence of Stochastic Processes*. Springer-Verlag, 1984.
- [38] N. U. Prabhu. *Stochastic processes: basic theory and its applications*. World Scientific, 2007.
- [39] J. C. Principe. *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. Springer, 2010.
- [40] N. L. Johnson S. Kotz. *Breakthroughs in Statistics: Foundations and basic theory*. Springer, 1992.

- [41] H. Scheffé. *The Analysis of Variance*. John Wiley and Harvard University Press Sons, Inc., 1959.
- [42] A. N. Shiryaev. *Probability*. Springer, 1996.
- [43] M. Hitsuda T. Hida. *Gaussian Processes*. American Mathematical Society, 1993.
- [44] D. A. Darling T. W. Anderson. Asymptotic theory of certain.
- [45] K. Takezawa. *Introduction to nonparametric regression*. John Wiley and Sons, 2006.
- [46] H. C. Thode. *Testing for Normality*. Marcel Dekker, 2002.
- [47] H. M. Rosenblatt W. T. Shearer et al. Lymphocyte subsets in healthy children from birth through 18 years of age. *Journal of Allergy and Clinical Immunology*, 112:973–980, 2003.
- [48] S. Wilks. *Mathematical Statistics*. Wiley, 1962.
- [49] C. Loader W.S. Cleveland. Smoothing by local regression: Principles and methods. Technical report, Bell Laboratories, 1996.
- [50] Y. Pawitan Y. Lee, j. Nelder. *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*. Chapman and Hall, 2006.
- [51] F. E. Young. The role of the fda in the effort against aids. *Public Health Reports*, 103(3):242–245, 1988.
- [52] G. Zyskind. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *The Annals of Mathematical Statistics*, 38(4):1092–1109, August 1967.