

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA
IN STATISTICA POPOLAZIONE E SOCIETÀ

**Teoria moderna della verosimiglianza:
applicazioni nella fisica delle particelle.**

Relatore:
PROF.SSA
ALESSANDRA SALVAN

Laureanda:
SILVIA CALDERAZZO

Anno Accademico 2009/2010

Indice

Introduzione	I
1 Fisica e statistica	1
1.1 Qualche cenno alla fisica delle particelle	1
1.2 Di come la statistica entrò nella ricerca fisica	6
1.3 Il <i>Banff Workshop</i>	10
1.3.1 Obiettivi	10
1.3.2 Considerazioni finali	11
1.4 Alcuni temi fisico-statistici	12
1.4.1 Scelta del livello di significatività	12
1.4.2 Scelta del paradigma statistico	14
1.4.3 Analisi cieche	15
2 Il modello statistico	19
2.1 Presentazione del modello	19
2.2 Richiami teorici	21
2.2.1 Funzione di verosimiglianza e stima di massima verosimiglianza	21
2.2.2 Intervalli di confidenza	23
2.2.3 Log-verosimiglianza profilo e risultati asintotici connessi	26
2.3 Applicazione al caso in esame	27
2.3.1 Modello	27
2.3.2 β noto	27
2.3.3 Una spiegazione elementare del problema	29

2.3.4	Presenza di parametri di disturbo	30
3	Significance functions	33
3.1	Un diverso modo di pensare	33
3.1.1	Un esempio: la media di una distribuzione normale . .	35
3.2	Applicazione al caso in esame	38
3.2.1	β, γ noti	38
3.2.2	Presenza di parametri di disturbo	50
	Conclusioni	55
	Bibliografia	57

Introduzione

Questa tesi si pone l'obiettivo di trattare, nei suoi aspetti qualitativi e quantitativi, un problema di origine fisica, che trova però una soluzione operativa nell'impiego di alcune particolari tecniche statistiche. Si presenteranno, con alcune rielaborazioni personali, principalmente i risultati presenti in Fraser, Reid e Wong (2004) e Davison e Sartori (2008).

Nel primo capitolo si cercherà di descrivere gli aspetti fisici e qualitativi del problema, ponendo inoltre l'accento su alcuni temi connessi alla recente introduzione della statistica nella ricerca fisica: la trattazione non ha in alcun modo la pretesa di risultare esuastiva, ha il solo scopo di introdurre l'“ordine di idee” nel solco del quale il problema ha avuto origine. I due linguaggi, fisico e statistico, risultano infatti per molti aspetti profondamente diversi, ed il tentativo di unificazione frequentemente non è stato indolore. Da un lato, infatti, c'è una visione fortemente deterministica o, quantomeno, univoca, della realtà, la visione fisica, dall'altra c'è la presenza dell'incertezza, la necessità di compiere scelte per le quali non esiste un criterio universale o algoritmico: scelte diverse, spesso tutte equamente valide, possono produrre risultati differenti. Questo, ai fisici, tendenzialmente non piace. Sarà quindi trattato il modo in cui queste due visioni siano state, in un certo senso, conciliate.

Nel secondo capitolo verrà presentato il modello statistico adottato, e verrà introdotta una descrizione più formale e rigorosa del problema.

Nel terzo capitolo si tratterà infine una possibile soluzione proposta di recente (Fraser, Reid e Wong, 2004 e Davison e Sartori, 2008), ponendo

l'accento sulla sua traduzione operativa e sulla sua applicazione al problema oggetto di analisi.

Raccogliere informazioni, comprendere diverse metodologie, sentir parlare in prima persona, attraverso i loro scritti, gli scienziati e gli statistici, si è rivelato, al di là delle difficoltà spesso incontrate, soprattutto un divertimento. L'auspicio è di essere riuscita, almeno in parte, a trasmettere questo piacere a quei pochi che leggeranno quanto scritto nelle prossime pagine.

Capitolo 1

Fisica e statistica

*“There are more things in heaven and earth, Horatio
than are dreamt of in your philosophy.”*

W. Shakespeare, Hamlet, Act I

1.1 Qualche cenno alla fisica delle particelle

Tutto nasce dall’infinitamente piccolo. Il problema statistico che verrà affrontato e discusso nel seguito deriva da un problema di natura fisica: la ricerca di particelle in grado di spiegare la struttura profonda dell’universo.

L’idea che tutta la materia sia formata da costituenti ultimi ed elementari che, tra loro combinati, danno vita alla molteplicità degli oggetti animati ed inanimati, ha le sue radici nell’antica Grecia, epoca nella quale il filosofo Democrito teorizzò l’esistenza di tali costituenti dando loro il nome di atomi, etimologicamente, entità indivisibili.

Per svariati secoli, la conoscenza degli atomi non oltrepassò la pura speculazione filosofica cui era stato dato inizio nell’epoca classica. Fu tra il XVII ed il XIX secolo che lo sviluppo delle scienze chimiche portò ad identificarne un numero relativamente ristretto (92): dalla loro combinazione si riteneva fossero formate le molecole e, dunque, tutte le sostanze presenti in natura. Benché tale concezione fosse tendenzialmente corretta, nessun progresso era

stato fatto nel comprendere la natura profonda degli atomi e i motivi fisici per cui essi differivano tra loro.

Il punto di svolta si ebbe sul finire del XIX secolo, quando la scoperta dei raggi catodici portò a ritenere che gli atomi contenessero corpuscoli molto leggeri, puntiformi e di carica negativa (gli elettroni). Essendo però l'atomo elettricamente neutro, si rivelò necessario ipotizzare l'esistenza di un componente di carica positiva, responsabile inoltre di gran parte della massa complessiva. In tale contesto, si fece strada il modello 'a panettone' di Thomson, nel quale si immaginava che l'atomo fosse una sorta, appunto, di panettone di carica positiva nel quale erano disseminati gli elettroni 'come l'uvetta'.

Nel 1911 Lord Rutherford rivelò l'inesattezza di tale teoria bombardando con *particelle alfa* (nuclei di elio, quindi carichi positivamente) sottili lamine d'oro. L'esperimento mostrò che parte di queste particelle era non solo deviata nella sua traiettoria, ma poteva essere addirittura respinta. Come scrisse lo stesso Lord Rutherford (Andrade, 1964): *"It was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you."* Lo studio del numero di particelle deviate e degli angoli di deviazione portò a teorizzare l'esistenza e le dimensioni di quello che venne definito il nucleo atomico.

Nel 1919 lo stesso Lord Rutherford compì la prima trasmutazione: trasformò il nucleo stabile di un elemento chimico, nel nucleo stabile di un elemento differente. Infatti, proiettando fasci di particelle alfa attraverso una strato di gas azoto, si accorse che, talvolta, il nucleo di quest'ultimo elemento catturava una particella alfa e si trasformava in un nucleo di ossigeno, emettendo una particella diversa da quella incidente. Questa, aveva carica esattamente opposta a quella dell'elettrone: era stata identificata la componente positiva dell'atomo, il protone.

Se la carica elettrica neutra dell'atomo era stata così spiegata, ancora restava da individuarne parte della massa, ed il motivo per il quale le cariche

elettriche positive del nucleo non si respingessero tra loro, ma dessero al contrario vita a nuclei stabili. Fu nel 1932 che Chadwick, bombardando con particelle alfa il berillio, scoprì che i nuclei di quest'ultimo elemento si disintegravano, emettendo una nuova particella sino ad allora sconosciuta, di carica nulla e di massa poco superiore a quella del protone: il neutrone.

Negli anni successivi, centinaia di altre particelle sono state rilevate, e molte altre sono state teorizzate per spiegare i fenomeni fisici osservati. Attualmente si ritiene che esistano dodici entità fondamentali (prive cioè di una sottostruttura), suddivise in due categorie che ne contengono sei ciascuna: leptoni e quark.

Leptoni e quark sono suddivisi in tre generazioni di massa crescente e differenziati, inoltre, sulla base della loro carica elettrica. La materia stabile dell'universo è costituita da particelle della prima generazione, in quanto le particelle più pesanti (appartenenti alle generazioni successive) tendono a decadere nel più stabile livello successivo.

L'esistenza dei quark non è stata ad oggi osservata direttamente: si ipotizza infatti che essi non possano esistere se non in gruppi di due o tre, essendo dotati di carica elettrica frazionaria (rispetto a quella dell'elettrone, assunta come riferimento). Tuttavia, la loro esistenza risulta necessaria per la formazione di particelle più complesse, come i protoni, i neutroni ed i mesoni (nel complesso, queste ultime tre particelle sono chiamate adroni).

Tra i leptoni, a ciascuna delle tre particelle dotata di massa e carica elettrica è associato un elemento della medesima generazione, di carica nulla e massa molto ridotta: i neutrini. L'elettrone stesso, leptone della prima generazione di carica negativa, ha associato il corrispondente 'elettrone-neutrino'.

Un quadro complessivo di quark e leptoni, delle rispettive masse e cariche, si può osservare in Tabella 1.1.

Il Modello Standard, cioè la teoria attualmente utilizzata per la spiegazione dei fenomeni atomici e sub-atomici, teorizza l'esistenza, e spiega matematicamente le interazioni, di tre forze fondamentali: la forza forte, la forza debole e la forza elettromagnetica. Queste hanno intensità e raggi differenti

Tipo di particella	Carica	Generazione (Massa)		
		1	2	3
Quark	+2/3	u (0.3)	c (1.5)	t (175)
Quark	-1/3	d (0.3)	s (0.5)	b (5)
Leptoni	-1	elettrone (5×10^{-4})	μ (0.1)	τ (1.8)
Leptoni (neutrini)	0	$\nu_e (< 3 \times 10^{-9})$	$\nu_\mu (< 2 \times 10^{-4})$	$\nu_\tau (< 0.02)$

Masses espresse in GeV/c^2 , la massa del protone in questa unità di misura è 0.9.

Tabella 1.1: Le particelle elementari

ed è inoltre noto che derivano dallo scambio di particolari particelle mediatrici, dette bosoni. Queste ultime sono: il fotone γ , mediatore della forza elettromagnetica, i bosoni W e Z, mediatori della forza debole (responsabile del decadimento radioattivo), ed il gluone g, mediatore della forza forte (che mantiene unito il nucleo).

Esiste tuttavia un'altra forza fondamentale che, ad oggi, non è stato possibile conciliare con il Modello Standard: la forza gravitazionale. Si ritiene che anch'essa derivi da una particella mediatrice, il gravitone, e l'attenzione della ricerca è ora concentrata sulla sua rilevazione, difficile a causa della debole intensità della forza. Il motivo per cui il Modello Standard è tuttora applicato e si rivela particolarmente appropriato nello studio delle particelle, è che la forza gravitazionale, a livello atomico, produce effetti così deboli da risultare trascurabili. Purtroppo non è così a livello macroscopico, dove essa agisce in modo tutt'altro che trascurabile, determinando, ad esempio, le orbite dei pianeti e molti altri fenomeni di esperienza comune. Si è ora alla ricerca di una teoria che concili le leggi dell'infinitamente piccolo e dell'infinitamente grande, in sostanza, una 'teoria del tutto' che fornisca un disegno complessivo della realtà.

Un altro ‘tassello mancante’ su cui le ricerche sono ora concentrate è il bosone di Higgs, ritenuto responsabile della massa di tutte le altre particelle: si suppone infatti che, immediatamente dopo il Big Bang, tutte le particelle fossero prive di massa. Non appena la temperatura scese sotto un determinato livello critico, si creò un campo di forza detto, appunto, ‘campo di Higgs’, e l’associato bosone. Qualunque particella interagisse con tale campo, attraverso il bosone di Higgs veniva dotata di massa. La sua scoperta porterebbe all’unificazione della forza debole e di quella elettromagnetica.

Allo scopo di confermare o confutare le teorie esistenti, diversi esperimenti sono stati e saranno condotti in noti centri di ricerca internazionale, come il CERN di Ginevra o il Fermilab di Chicago, dove sofisticati ed imponenti macchinari, tecnicamente chiamati acceleratori di particelle, proiettano e fanno collidere fasci di protoni, rilevandone poi gli esiti. Si tratta di tunnel circolari la cui lunghezza può raggiungere svariati chilometri (l’acceleratore del CERN, l’LHC, ha una circonferenza di 27 Km), i protoni sono proiettati in direzioni opposte e collidono al centro del percorso, dove è posizionato un rivelatore che raccoglie informazioni relative agli esiti degli scontri: in seguito alle collisioni, infatti, si creano nuove particelle, trasformando così l’energia cinetica presente in massa. I rivelatori osservano la traccia delle nuove particelle, la curvatura del campo magnetico ed il momento, fornendo inoltre informazioni sull’identità della particella stessa.

Dopo il guasto avvenuto nel Settembre 2008, l’LHC (Large Hadron Collider) di Ginevra alle 13:06 del 30 Marzo 2010 ha ripreso a funzionare alla massima energia (7 TeV), dando luogo alla prima collisione. Nei prossimi mesi una mole elevatissima di dati sarà raccolta: solo allora si potrà, forse, sapere se i modelli finora noti siano validi, o se la realtà, ancora una volta, rivelerà qualcosa di completamente nuovo rispetto alle conoscenze esistenti.

Riferimenti ai temi introdotti si possono ritrovare in Penrose (2005, Capitolo 25), Lyons (2008) e all’indirizzo www.cern.ch.

1.2 Di come la statistica entrò nella ricerca fisica

La domanda sorge spontanea: in tutto questo, cosa c'entra la statistica?

La risposta è tutt'altro che banale e, per comprendere il ruolo della statistica e del calcolo delle probabilità nella ricerca fisica, è necessario fare qualche cenno a come, nel corso dei secoli, gli scienziati si siano avvicinati alla conoscenza della realtà.

Esistono due fondamentali processi conoscitivi che possono portare alla formulazione di nuove leggi scientifiche: il processo induttivo ed il processo deduttivo.

L'induzione ha il suo punto di partenza nelle osservazioni empiriche particolari, dalle quali vengono astratte (indotte) le leggi universali, le leggi, cioè, universalmente valide all'interno del loro campo di applicabilità, che può avere limiti spaziali, temporali, legati alle proprietà fisico-chimiche degli elementi coinvolti ... etc. L'obiettivo finale è la ricerca di leggi dotate di un campo di applicabilità che sia il più esteso possibile, esso è infatti indice della generalizzabilità (e dunque dell'importanza) della legge scoperta. L'induzione è anche definita "un'inferenza ampliativa ma incerta", questo perché, partendo dal particolare (necessariamente imperfetto), per giungere all'universale, l'unica via percorribile per ottenere la certezza sarebbe l'osservazione di tutti i possibili elementi del particolare, virtualmente infiniti.

La deduzione, al contrario, è detta "inferenza non ampliativa, ma necessaria": il punto di partenza non è più la realtà, ma una serie di postulati indimostrabili, dai quali sono dedotte via via le leggi scientifiche attraverso dimostrazioni consequenziali e necessarie. Anche in questo caso esiste un campo di applicabilità, ma è determinato senza la necessità di un confronto diretto con la realtà, rappresentando anch'esso una conseguenza necessaria di premesse date.

La nascita del moderno metodo scientifico si deve a Galileo Galilei che agli inizi del XVII secolo ne diede una formulazione rigorosa. Tale metodo è

insieme induttivo e deduttivo, partendo infatti dall'osservazione della realtà, giunge all'elaborazione di una legge che viene sottoposta ad un'analisi matematica rigorosa, necessaria alla formulazione delle ipotesi sperimentali. Il processo si conclude con la verifica empirica della legge attraverso le ipotesi formulate.

Le idee principali alla base dello studio scientifico dei fenomeni sono ben descritte nella “Lettera a Madama Cristina di Lorena” (Galiei, 1615):

“[...] Essendo la natura inesorabile ed immutabile, e mai non trascendente i termini delle leggi impostegli, come quella che nulla cura che le sue recondite ragioni e modi d'operare sieno o non sieno esposti alla capacità degli uomini; pare che quello degli effetti naturali che o la sensata esperienza ci pone dinanzi a gli occhi o le necessarie dimostrazioni ci concludono, non debba in conto alcuno esser revocato in dubbio.”

In altre parole, alla base della conoscenza del mondo naturale, vi era la concezione che la realtà fosse regolata da leggi deterministiche e che queste potessero essere individuate attraverso la “sensata esperienza” (l'esperimento scientifico) e le “necessarie dimostrazioni” (l'analisi matematica dei risultati forniti dall'osservazione della realtà). Il libro della natura era dunque scritto in linguaggio matematico: il compito dello scienziato era comprendere tale linguaggio e le necessarie relazioni causali che la natura, attraverso di esso, manifestava. È la nascita della concezione meccanicistica della realtà.

Dall'epoca di Galileo, la conoscenza del mondo fisico fu notevolmente ampliata ed approfondita, ma l'assunzione che l'oggetto della ricerca, la realtà fisica, fosse regolata da leggi certe ed inesorabili non venne messa in discussione sino a metà del XIX secolo.

La svolta decisiva per il passaggio a leggi di tipo probabilistico nella fisica si ebbe con Ludwig Boltzmann (si veda per ulteriori riferimenti, Boltzmann, 2010). Egli rifiutò, sostanzialmente, la visione settecentesca secondo la quale i fenomeni termici potessero essere ricondotti alle proprietà di un fluido detto

“calorico”, affermando e dimostrando che essi sono piuttosto dovuti all’interazione reciproca tra atomi e molecole. Esaminando lo stato gassoso, nel quale essi sono completamente liberi di muoversi, formulò la teoria cinetica dei gas. Tralasciando una descrizione più accurata di tale teoria, lontana dagli scopi di questa introduzione, la svolta fondamentale si ebbe grazie al fatto che lo studio dell’evoluzione di atomi e molecole presi singolarmente, data la loro numerosità, era del tutto impensabile: con l’introduzione di valori medi e proprietà asintotiche, era nata la prima legge fisica probabilistica e la meccanica statistica.

La vera rivoluzione si basò sul fatto che il calcolo delle probabilità (e dunque la statistica) non solo sopperiva alla mancanza di parte delle possibili osservazioni nella fase induttiva del processo, ma entrava anche a far parte integrante di quella deduttiva, rappresentando un metodo valido per la formulazione di una legge a partire da premesse date.

Si è a questo punto facilmente indotti a pensare che venne accettata, con tali teorie, l’idea che il caso fosse in qualche modo insito nella natura stessa e che non fosse più possibile spiegare i fenomeni fisici in modo totalmente meccanicistico. Ciò non è completamente vero: sebbene la meccanica statistica faccia uso del calcolo delle probabilità, anch’esso è sempre e comunque regolato da leggi immutabili e matematiche, come enfatizzò lo stesso Boltzmann in un articolo del 1872 (Boltzmann, 2010):

“La determinazione dei valori medi è compito del calcolo delle probabilità. Ma sarebbe un errore credere che la teoria del calore sia per questo soggetta all’incertezza, perché ivi vengono utilizzati i teoremi del calcolo delle probabilità. Non si confonda una legge dimostrata in modo incompleto, la cui correttezza in seguito a ciò è problematica, con una legge perfettamente dimostrata del calcolo delle probabilità; quest’ultima rappresenta, come il risultato di qualsiasi altro calcolo, una conseguenza necessaria di premesse date e si rivela altrettanto fondata nell’esperienza, non appena queste siano corrette, se solo viene portato all’osservazione un

numero sufficientemente grande di casi, cosa che avviene sempre nella teoria del calore visto l'enorme numero di molecole."

Qualora si presenti un numero sufficientemente elevato di casi, si ha quindi solo un passaggio ad un nuovo determinismo, regolato dalle leggi del calcolo delle probabilità.

Può essere che tale 'determinismo probabilistico' sia valido perché, effettivamente, la realtà è dominata in ultima istanza da leggi casuali, o può essere che lo sia perché sopperisce adeguatamente (in modo matematico e rigoroso) alla finitezza delle osservazioni. In ogni caso, grazie all'opera di Boltzmann, è stata sancita la sua validità nella formulazione di leggi fisiche.

La portata di simili affermazioni in un contesto dominato dalla ricerca di leggi certe, era comunque notevole. Ne sorse un dibattito che proseguì nel corso dei decenni successivi, lo stesso Einstein vi partecipò e a tal proposito pronunciò le famose parole: *"Dio non gioca a dadi col mondo"*, sostenendo così che, per quanto nostre le misurazioni possano essere imperfette, il principio regolatore dell'universo rimane un principio certo e non casuale.

Forse, il contrasto tra la visione deterministica e quella probabilistica è destinato a non trovare una soluzione definitiva, anche perché, attiene, almeno in parte, più alla metafisica che alle scienze fisiche modernamente intese.

Tuttavia, sia che la realtà fisica sia dominata da leggi deterministiche, sia che vi sia una residua, ineluttabile, casualità, la statistica ha assunto negli ultimi decenni un ruolo sempre più importante nella ricerca scientifica: le moderne tecnologie, in grado di produrre un'elevata mole di dati, la difficoltà nell'osservazione di oggetti infinitesimi, il costo notevole dei singoli esperimenti, hanno giocato ulteriormente in suo favore. Modelli statistici si sono resi necessari per la descrizione della realtà di interesse, e varianze ed intervalli di confidenza hanno spesso giocato un ruolo decisivo nell'accettare o confutare ipotesi di ricerca. Le parole di Lord Rutherford *"If your experiment needs statistics, you ought to have done a better experiment"*, sembrano ormai, definitivamente, superate.

1.3 Il *Banff Workshop*

In questo contesto, sempre più frequentemente negli ultimi anni, sono stati organizzati congressi internazionali, all'interno dei quali fisici e statistici hanno discusso e collaborato, tanto per risolvere problemi sorti dall'interpretazione quantitativa di dati prodotti da recenti ricerche, quanto in via preliminare a nuovi esperimenti. Ne è un esempio l'incontro avvenuto nel luglio 2006 presso The Banff International Research Station for Mathematical Innovation and Discovery (BIRS) (per il report finale si veda Linnemann, Lyons e Reid, 2006). Pur non essendo l'unica occasione in cui conoscenze fisiche e statistiche si siano confrontate ed unite, su questo congresso verrà qui posta una particolare attenzione in quanto costituirà il riferimento per parte dei dati e dei risultati discussi nel seguito.

1.3.1 Obiettivi

Gli obiettivi statistici specifici dell'incontro di Banff sono stati così riassunti:

1. La costruzione di intervalli di confidenza per un parametro di interesse in presenza di parametri di disturbo;
2. Soluzioni a problemi di classificazione: separazione di eventi di interesse da eventi di disturbo;
3. Determinazione della bontà del modello stimato.

I partecipanti, 33 in totale tra fisici, astrofisici e statistici, sono stati chiamati a confrontare diversi approcci per la risoluzione dei problemi sopra elencati, confrontando quindi i risultati ottenuti. Parallelamente, attraverso una serie di interventi, i fisici hanno esposto i principali argomenti connessi alle ricerche oggetto di studio, e gli statistici i principali approcci utilizzati e le nuove tecniche scoperte, illustrando i miglioramenti che potrebbero essere introdotti negli usuali metodi di analisi.

Benché nel corso dell'incontro numerosi temi di interesse siano stati toccati, l'attenzione si soffermerà qui sul primo obiettivo: la costruzione di intervalli di confidenza in presenza di parametri di disturbo. La presenza di eventi di disturbo indistinguibili dagli eventi di interesse, infatti, crea delle difficoltà nel comprendere (probabilisticamente) se l'evento di interesse si sia effettivamente verificato o meno: si ponga, ad esempio, che l'evento di interesse sia la presenza di segnali derivanti da nuove particelle, e si ipotizzi di essere a conoscenza della stima (o del valore esatto) del numero di segnali ricevuti, contemporaneamente, da particelle di disturbo e di interesse, tra loro indistinguibili. Il numero di segnali di interesse ricevuti non può essere inferiore a zero, dunque il totale dei segnali è vincolato ad assumere valori uguali o maggiori di quelli stimati per le particelle di disturbo. Il livello di confidenza espresso deve quindi tenere conto di questo vincolo ed è di difficile determinazione secondo i metodi usuali. Su questo punto in particolare verteranno i capitoli seguenti.

1.3.2 Considerazioni finali

Tralasciando momentaneamente i risultati più strettamente teorici e scientifici, può essere interessante (e divertente) capire quali siano state le conclusioni, in un certo senso umane, di questi incontri tra studiosi della casualità e della materia. Citando Linnemann, Lyons e Reid (2006): *“It was noted, however, that while Astrophysics has been successful in involving statisticians in their analyses to the extent where their names appear on experimental papers, this is usually not the case in particle physics. Several reasons have been put forward to explain this. One is that statisticians like analysing real data, with all its interesting problems. But particle physics experimental collaborations tend to be very jealous about their data, and are unwilling to share it with anyone outside the collaboration until it is too old to be interesting. This results in particle physicists asking statisticians only very general questions, which the statisticians regard as unchallenging and boring.”*

Sembra sostanzialmente esistere ancora qualche riluttanza nell'ammet-

tere a pieno titolo gli statistici nello studio della fisica delle particelle. Va detto, che l'introduzione della statistica in questo ambito è relativamente recente. L'auspicio è che nel futuro si riesca a guardare alla ricerca ed alla conoscenza della realtà a prescindere da pregiudizi e scissioni tra le discipline, considerando che, in fondo, l'obiettivo è uno, e comune.

1.4 Alcuni temi fisico-statistici

Nei prossimi paragrafi si cercherà di fornire qualche cenno ad alcune tematiche che hanno interessato la ricerca fisica (e non solo) successivamente all'introduzione della statistica. La presentazione proposta non ha alcuna pretesa di fornire una trattazione esaustiva né degli argomenti introdotti, né, tantomeno, della totalità dei problemi che la progettazione di un esperimento e l'analisi dei risultati porta con sé. Ha il solo proposito di presentare in modo molto generale alcune tematiche che il problema che verrà discusso nel seguito tocca, in modo più o meno rilevante.

1.4.1 Scelta del livello di significatività

È lecito chiedersi perché i livelli di significatività ricoprano un ruolo così importante nelle ricerche condotte: la domanda può apparire banale, ma alla base vi è una lunga discussione su come debbano essere impiegati nell'acclarare o meno una nuova scoperta. L'introduzione della statistica, infatti, porta con sé un nuovo tipo di decisione: la scelta della probabilità con la quale si è disposti a rifiutare l'ipotesi nulla (che non introduce nuove entità teoriche) quando questa è vera, in altre parole, quello che viene tecnicamente chiamato il livello di significatività α . Fissando un α troppo elevato, si corre il rischio di rifiutare troppo facilmente l'ipotesi nulla; al contrario, fissando un α troppo ridotto, il rischio è di giungere con scarsissime probabilità a rifiutare l'ipotesi nulla (anche quando è falsa) e ad annunciare nuove scoperte. La seconda condizione si verifica in quanto, diminuendo α , aumenta la probabilità di errore di secondo tipo, generalmente indicata con β : quest'ultima

rappresenta, appunto, la probabilità di rifiutare l'ipotesi alternativa quando è vera (o, analogamente, accettare l'ipotesi nulla quando è falsa). Sulla base di queste due probabilità, inoltre, si è costruita tutta una parte della teoria statistica legata alle procedure ottime che sono, per definizione, quelle procedure che consentono di minimizzare β a parità di α fissato.

È auspicabile che il test sia non distorto, ossia che α risulti inferiore ad $1 - \beta$ (cioè che sia più facile rifiutare l'ipotesi nulla quando è falsa piuttosto che quando è vera), così come è auspicabile che la procedura utilizzata sia una procedura ottima. Volendo, inoltre, attuare un meccanismo conservativo nei confronti dell'ipotesi nulla, sarebbe opportuno richiedere α inferiore a β . Tuttavia, poste queste condizioni (quando possibile), la scelta di α resta a discrezione del ricercatore.

Nell'ambito degli esperimenti fisici è stato osservato (Linnemann, Lyons e Reid, 2006) che α pari a 0.05 o 0.001 hanno portato ad affermare la presenza di effetti che si sono in seguito rivelati inesistenti. La linea guida attualmente seguita suggerisce l'utilizzo di un α pari a circa $1 - \Phi(5)$, con Φ funzione di ripartizione della normale standard (numericamente: $2,86 \times 10^{-7}$), ossia una probabilità inferiore ad uno su un milione. Come discusso nell'intervento di David van Dyk e Luc Demortier (Linnemann, Lyons e Reid, 2006), il dubbio che una probabilità così bassa porti ad ignorare effetti sistematici di una certa rilevanza, è presente. Inoltre, scegliere a priori un livello di significatività utilizzabile in tutti gli esperimenti, indipendentemente dalle ipotesi poste sotto verifica e dalla numerosità campionaria, può risultare pericoloso. Ciononostante, come scrive Berger (2008): *“In practice, however, declaration of a discovery often does have other effects - e.g., people stop research along lines that are incompatible with the discovery - so there is a serious cost to erroneous claims of discovery (in addition to having to return the Nobel prizes), even if there is a possibility of later correction.”*

1.4.2 Scelta del paradigma statistico

Un altro tema sul quale è ancora vivo il dibattito, è la scelta del paradigma statistico da utilizzare nell'analisi dei dati: la contrapposizione riguarda il paradigma bayesiano e quello fisheriano-frequentista. Non si tratta, come potrebbe apparire, solo di una divergenza sulle tecniche più efficienti, ma ha alla base un'assunzione, che potremmo dire filosofico-epistemologica, su come debba avvenire la conoscenza della realtà.

L'approccio bayesiano (si veda per ulteriori riferimenti Liseo, 2008, Capitolo 3) sfrutta per l'inferenza un teorema del calcolo delle probabilità (noto, appunto, come teorema di Bayes) il quale, nella sua forma più semplice, afferma che, dati due eventi F ed E ,

$$P(F | E) = \frac{P(F)P(E | F)}{P(E)}. \quad (1.1)$$

Considerando F il parametro di interesse dell'esperimento ed E il risultato sperimentale osservato (il vettore $y = (y_1, \dots, y_n)$), si ha

$$\pi(\theta | y) = \frac{\pi(\theta)L(\theta; y)}{\int_{\Theta} \pi(\theta)L(\theta; y)d\theta}. \quad (1.2)$$

La densità del parametro, dati gli eventi osservati, è detta probabilità *a posteriori* ed è una misura che deriva non solo dall'informazione fornita dall'esperimento, la verosimiglianza $L(\theta; y)$, ma anche da una probabilità $\pi(\theta)$ assegnata al parametro sulla base delle conoscenze dello sperimentatore: essa è detta, al contrario, probabilità *a priori*. La sua scelta, essendo discrezionale, porta ad accusare questo approccio di soggettività: due scienziati che scelgano due *a priori* diverse, infatti, otterranno risultati differenti. Per ovviare a questo problema, si utilizzano le cosiddette distribuzioni a priori non informative. Come affermato da Liseo (2008, Capitolo 5), tuttavia: “*Non è possibile dare una definizione soddisfacente del concetto di distribuzione non informativa, proprio perché è altrettanto sfuggente il concetto di informazione in senso probabilistico. Dal punto di vista operativo, una distribuzione non informativa dovrebbe concretizzarsi in una distribuzione iniziale basa-*

ta non già sulle informazioni a disposizione del ricercatore, bensì costruita in modo convenzionale, tenendo conto solamente della struttura statistica dell'esperimento."

Il paradigma frequentista, al contrario, si basa sulla concezione che nell'analisi dei dati osservati, essi debbano 'parlare da sé' fornendo conferme o rettifiche alle ipotesi del ricercatore senza che quest'ultimo sia in alcun modo influenzato dal proprio 'punto di vista'. A partire da questo presupposto, si sono sviluppati i diversi risultati basati, in larga misura, sulla funzione di verosimiglianza, atti ad individuare i livelli di significatività osservati e ad attuare la verifica di ipotesi.

I ricercatori fisici, stando a quanto affermato da Lyons (2008), sono tendenzialmente più inclini ad adottare un approccio frequentista, temendo forse, in caso contrario, affermazioni simili a quelle dell'*ipse dixit* pre-galileiano, la cui validità era determinata più dall'autorevolezza di chi le pronunciava, che non da razionali e riproducibili osservazioni empiriche.

La scelta dell'uno o dell'altro paradigma non è tuttavia così esclusiva o categorica come potrebbe sembrare: sono stati infatti recentemente sviluppati approcci che prevedono una commistione dei due, come il metodo bayesiano empirico, nel quale la scelta dell'*a priori* non avviene indipendentemente dalle informazioni raccolte, bensì esso è stimato in modo frequentista sulla base dei dati osservati (si veda Efron, 2005). Inoltre, risultati sperimentali (come il caso oggetto di analisi) hanno dimostrato che, al variare degli scopi dell'analisi, l'uno o l'altro metodo può risultare di volta in volta più preciso, indebolendo così l'eventuale propensione ad abbracciare 'ciecamente' l'una o l'altra metodologia.

1.4.3 Analisi cieche

Pur abbracciando un approccio frequentista, il problema della soggettività non è totalmente risolto: la progettazione dell'esperimento, la scelta di α prima menzionata, la formulazione delle ipotesi, la raccolta dei dati e, soprattutto, la formulazione del modello opportuno, possono introdurre ulte-

riori distorsioni. Come scrisse infatti Dunnington in un articolo del 1932 nel quale presentava i risultati delle proprie ricerche (Dunnington, 1932):

“It is also desirable to emphasize the importance of the human equation in accurate measurements such as these. It is easier than is generally realized to unconsciously work toward a certain value. One cannot, of course, alter or change natural phenomena...but one can, for instance, seek for those corrections and refinements which shift the results in the desired direction.”

È infatti possibile che lo sperimentatore attui inconsciamente dei meccanismi di distorsione volti a verificare le proprie ipotesi o a confermare le proprie convinzioni. In Klein e Roodmann (2005) è riportato un esempio relativo alla stima di alcuni specifici parametri da parte del PDG (Particle Data Group): dalle pubblicazioni annuali contenenti gli esiti delle misurazioni, si osserva che esiste un’associazione (in termini di χ^2) più elevata delle misurazioni ottenute di anno in anno con i risultati precedenti, piuttosto che con la stima finale oggi accettata. Come sottolineato, questo non costituisce una dimostrazione certa della presenza del fenomeno di distorsione, ma, unitamente ad analoghi casi e considerazioni, ha messo in guardia sulla sua possibilità, ed ha portato a sviluppare tecniche che permettessero di eliminarlo almeno in parte.

Sono state introdotte a questo scopo le cosiddette ‘analisi cieche’ (si veda per un riferimento completo, Lyons, 2008 e Klein e Roodmann, 2005): esse comprendono diverse procedure, nelle quali l’idea di fondo è che la scelta del modello opportuno e della procedura di analisi debba avvenire senza essere a conoscenza dei risultati reali, completi, dell’esperimento; queste procedure includono, tra le altre:

1. L’analisi preliminare solo di parte dei risultati (*hidden signal box*);
2. L’impiego di una caratteristica non nota nel corso dell’esperimento che influenza solo ed unicamente il risultato finale (*hidden answer*);

3. L'aggiunta o la sottrazione di valori casuali ai risultati dell'esperimento, rimossi solo dopo che ogni scelta metodologica è stata compiuta.

Per questo motivo, verificandosi nell'esperimento risultati incompatibili con il modello (come nel caso oggetto di analisi), l'idea di riformulare il modello a posteriori è accolta con scarso entusiasmo dai ricercatori fisici.

Forse, tuttavia, la prova più rassicurante del non totale condizionamento dei risultati ottenuti da parte dello sperimentatore è ancora fornita, come sosteneva Popper (1957), dalle falsificazioni delle teorie ipotizzate. Esse, infatti, *“ci rassicurano sul fatto che, benché le nostre teorie siano nostre stesse invenzioni, esse sono, nondimeno, genuine asserzioni sul mondo; possono infatti scontrarsi con qualcosa che non abbiamo mai fatto.”*

Capitolo 2

Il modello statistico

*“Essentially, all models are false,
but some are useful.”*

George E.P.Box

2.1 Presentazione del modello

Il problema, introdotto qualitativamente nel Capitolo 1, può essere così tradotto in termini statistici: si tratta di riassumere l'evidenza relativa alla rilevazione di un segnale, proveniente da alcune particolari particelle di interesse, in presenza di segnali di disturbo (derivanti da altre particelle), tenendo inoltre conto dell'efficienza degli strumenti di rilevazione.

La presenza del segnale di interesse può essere descritta come un arrivo casuale generato da un processo di Poisson. Si assume dunque che il numero di segnali rilevati in un singolo esperimento segua una legge di Poisson di media, incognita, pari al numero di arrivi atteso nel corso del singolo esperimento. Nel seguito verrà indicata come: $Poi(\mu)$.

Si assume che le particelle di disturbo si distribuiscano, nel numero dei loro arrivi casuali, come una $Poi(\beta)$. Verranno affrontate nel seguito le due possibili specificazioni del problema, con β assunto noto o da stimare.

Utilizzando la proprietà per cui la somma di due distribuzioni di Poisson

indipendenti è nuovamente una distribuzione di Poisson di media pari alla somma delle medie, è possibile descrivere il fenomeno oggetto di studio come la realizzazione di una $Poi(\beta + \mu)$, dove μ è assunto maggiore o uguale a zero.

Allo scopo di tenere conto dell'efficienza dello strumento di misurazione, è introdotto un nuovo parametro di disturbo che verrà indicato con γ . La distribuzione del numero di arrivi osservato risulta dunque, infine: $Poi(\gamma\mu + \beta)$. Nel seguito, per semplicità di notazione, si porrà $\theta = \gamma\mu + \beta$.

L'obiettivo è di individuare l'evidenza relativa a μ : alla luce dei dati raccolti, μ può essere supposta pari a zero o ci sono evidenze significative a favore dell'effettiva presenza del segnale? In termini di ipotesi questo si riduce a verificare:

$$H_0 : \theta = \beta \quad \text{contro} \quad H_1 : \theta > \beta$$

Un ulteriore, e in un certo senso analogo, passo consiste nella costruzione di intervalli di confidenza con livello $1 - \alpha$ fissato per μ .

Lo spazio parametrico naturale per il modello presentato è costituito dall'intervallo $(0, \infty)$, θ può dunque, teoricamente, assumere tutti i valori compresi in tale intervallo. Tuttavia, perché il modello sia fisicamente sensato, è ragionevole porre che μ (il numero di segnali di particelle di interesse ricevuti) sia non negativo. Se β è una costante nota strettamente maggiore di 0, questo implica una restrizione dei valori ammissibili per il parametro θ , che dovrà conseguentemente risultare maggiore o uguale a β .

Questa restrizione ha conseguenze molto rilevanti nella costruzione di intervalli di confidenza per μ , in quanto, in particolare in presenza di un numero di arrivi relativamente piccolo, tali intervalli possono essere in parte (o del tutto) al di fuori della regione $[\beta, \infty)$: in questi casi la probabilità (il livello di confidenza espresso) non ha apparentemente più senso. Il problema è stato introdotto in Mandelkern (2002) e trattato in diversi articoli, tra cui, Fraser, Reid e Wong (2004), Davison e Sartori (2008).

Per comprendere meglio l'affermazione fatta e per fissare la notazione, è opportuno introdurre alcuni concetti teorici.

2.2 Richiami teorici

Il riferimento a questo trattato nella seguente sezione può essere ritrovato in Pace e Salvan (2001, Capitoli 1 e 3).

Si assume che i dati osservati y^{oss} siano realizzazione di una variabile casuale Y dotata di una distribuzione di probabilità che si indicherà con P^0 . L'insieme di tutti i possibili valori che y può assumere è detto spazio campionario e verrà indicato con \mathcal{Y} .

L'individuazione di un modello statistico \mathcal{F} consiste nella specificazione di una famiglia di distribuzioni di probabilità compatibili con y^{oss} . Se P^0 appartiene ad \mathcal{F} , si dice che il modello è correttamente specificato.

P^0 può essere rappresentato utilizzando la funzione di probabilità, nel caso di variabili discrete, o densità, nel caso di variabili continue; tale funzione verrà indicata, in modo unificato, con $p(y)$.

Spesso, gli elementi di \mathcal{F} sono individuati attraverso un parametro θ , che rappresenta il parametro ignoto della funzione di densità di probabilità di Y . Si scriverà sinteticamente $\mathcal{F} = \{p(y; \theta), \theta \in \Theta\}$. Se la corrispondenza tra Θ e \mathcal{F} è biunivoca, allora il parametro θ è identificabile.

Inoltre, si dice che \mathcal{F} è un modello statistico parametrico se Θ è un sottoinsieme di \mathbb{R}^p , cioè se θ è un vettore di p costanti reali.

Se \mathcal{F} è correttamente specificato e se θ è identificabile, allora esiste un unico valore $\theta^0 \in \Theta$ tale che $p(y; \theta^0)$ è la densità di $P^0(y)$.

2.2.1 Funzione di verosimiglianza e stima di massima verosimiglianza

Si può a questo punto definire la funzione di verosimiglianza. Sia \mathcal{F} un modello statistico parametrico per i dati y con funzione di densità di probabilità $p(y; \theta)$, con $\theta = (\theta_1, \dots, \theta_p) \in \Theta \subseteq \mathbb{R}^p$. La funzione di verosimiglianza per θ basata sui dati y , $L : \Theta \rightarrow \mathbb{R}^+$ è definita da

$$L(\theta) = p(y; \theta)c(y)$$

funzione della sola variabile θ , con y fissato al valore osservato, in cui $c(y)$ rappresenta una costante moltiplicativa che non dipende da θ .

Frequentemente, $y = (y_1, \dots, y_n)$ è un vettore di osservazioni e si assume che esse siano realizzazioni indipendenti ed identicamente distribuite di (Y_1, \dots, Y_n) , repliche di una stessa variabile di base. In tal caso, indicata con $p_i(y_i, \theta)$ la funzione di probabilità di densità marginale di dell' i -esima variabile, la funzione di verosimiglianza assume la forma

$$L(\theta) = \prod_{i=1}^n p_i(y_i; \theta).$$

Per una semplificazione nella scrittura e nei calcoli, si preferisce generalmente utilizzare una sua trasformata monotona, detta funzione di log-verosimiglianza

$$l(\theta) = \log L(\theta) = \sum_{i=1}^n \log p_i(y_i; \theta)$$

dove, se $L(\theta) = 0$, viene assunta $l(\theta) = -\infty$.

Se $l(\theta)$ è differenziabile almeno tre volte, con derivate parziali continue in Θ , allora il modello è con verosimiglianza regolare ed è possibile definire due nuove quantità: la funzione score e la matrice di informazione osservata.

La funzione score è ottenuta dal vettore delle derivate parziali prime di $l(\theta)$

$$l_*(\theta) = \left(\frac{\partial l(\theta)}{\partial \theta_1}, \dots, \frac{\partial l(\theta)}{\partial \theta_p} \right).$$

La matrice di informazione osservata è data dall'opposto delle derivate seconde di $l(\theta)$

$$j(\theta) = \left[-\frac{\partial^2 l(\theta)}{\partial \theta_r \partial \theta_s} \right].$$

La stima di massima verosimiglianza di θ è, per definizione, un valore $\hat{\theta} \in \Theta$ tale che $L(\hat{\theta}) \geq L(\theta)$, $\theta \in \Theta$. Essendo $l(\theta)$ una trasformata monotona

di $L(\theta)$, la medesima relazione è valida per definire la stima di massima verosimiglianza sostituendo $l(\theta)$ ad $L(\theta)$.

Applicando le condizioni necessarie del primo ordine e sufficienti del secondo ordine per l'esistenza di massimi e minimi relativi, si ottiene una stima di massima verosimiglianza per θ ponendo $l_*(\theta) = 0$ qualora il segno di $l_{**}(\theta)$ risulti negativo (o la matrice di informazione osservata risulti definita negativa, nel caso multiparametrico); se la funzione di verosimiglianza tende a $-\infty$ agli estremi del dominio, il valore ottenuto è la stima di massima verosimiglianza per θ sui dati y .

Il significato dell'informazione osservata è ulteriormente esplicitato dal seguente risultato, ottenuto mediante lo sviluppo di Taylor di $l(\theta)$ centrato in $\hat{\theta}$ (caso monoparametrico):

$$l(\theta) = l(\hat{\theta}) + l_*(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2} l_{**}(\hat{\theta})(\theta - \hat{\theta})^2 + \dots \quad (2.1)$$

Se $\hat{\theta}$ è stima di massima verosimiglianza per θ , $l_*(\hat{\theta})$ risulta pari a zero, dunque $l(\theta)$ è approssimata da una parabola con vertice in $\hat{\theta}$ e curvatura dominata dall'informazione osservata: maggiore è la curvatura, più la verosimiglianza è concentrata intorno a $\hat{\theta}$.

2.2.2 Intervalli di confidenza

Per calcolare intervalli di confidenza e test per θ sono disponibili tre fondamentali risultati:

1. L'approssimazione asintotica della distribuzione dello stimatore alla distribuzione normale;
2. L'approssimazione al χ^2 per il log-rapporto di verosimiglianza;
3. L'approssimazione alla distribuzione normale della statistica *score*.

Il primo risultato afferma che, nei modelli statistici parametrici regolari, quando θ è il vero valore del parametro,

$$\hat{\theta} \sim N(\theta, j(\hat{\theta})^{-1}). \quad (2.2)$$

Un intervallo di confidenza per θ di livello α sarà dunque ottenuto mediante

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} j(\hat{\theta})^{-1/2}.$$

Viene inoltre introdotta la quantità

$$r_e(\theta) = (\hat{\theta} - \theta) \sqrt{j(\hat{\theta})}, \quad (2.3)$$

che ha distribuzione approssimata $N(0,1)$ sotto θ .

L'approssimazione risulta in genere più accurata, quanto più la forma della log-verosimiglianza normalizzata è ben approssimata da una parabola. Inoltre, all'aumentare dell'informazione osservata, a parità di α fissato, diminuisce l'ampiezza degli intervalli di confidenza. Gli intervalli ottenuti, tuttavia, non sono invarianti rispetto a riparametrizzazioni.

Nel caso multiparametrico, $\theta = (\theta_1, \dots, \theta_p)$ e si ha analogamente che

$$\hat{\theta} \sim N_p(\theta, j(\hat{\theta})^{-1}).$$

Un intervallo di confidenza di livello $1 - \alpha$ per la q -esima componente θ_q è dato da

$$\hat{\theta}_q \pm z_{1-\frac{\alpha}{2}} [j(\hat{\theta})^{-1/2}]_{q,q}.$$

dove $[j(\hat{\theta})^{-1/2}]_{q,q}$ rappresenta la radice della componente di posto (q, q) dell'inversa della matrice di informazione osservata.

Nei risultati richiamati $j(\hat{\theta})$ può essere sostituito da $i(\theta)$ o $i(\hat{\theta})$.

Il secondo risultato porta alla costruzione di intervalli di confidenza per θ basati direttamente sulla funzione di verosimiglianza; tali intervalli risultano invarianti rispetto a riparametrizzazioni e godono di una maggior precisione, soprattutto quando la funzione di verosimiglianza non è simmetrica e ben

approssimabile mediante una parabola. Si tratta di intervalli del tipo

$$\hat{\Theta}(Y) = \left\{ \theta \in \Theta : l(\theta) - l(\hat{\theta}) \geq -\frac{1}{2} \chi_{p;1-\alpha}^2 \right\},$$

dove p rappresenta la dimensione del vettore dei parametri θ .

Sono a questo punto definite le seguenti quantità:

$$W_e(\theta) = (r_e(\theta))^2 \quad (2.4)$$

$$W(\theta) = 2(l(\hat{\theta}) - l(\theta)) \quad (2.5)$$

$$r(\theta) = \text{sgn}(\hat{\theta} - \theta) \sqrt{W(\theta)} \quad (2.6)$$

$$s(\theta) = j(\hat{\theta})^{-1/2} \partial l(\theta) / \partial \theta. \quad (2.7)$$

$s(\theta)$ è la statistica *score* sopra citata, ed ha distribuzione approssimata, sotto θ , $N(0, 1)$.

Con $p = 1$, sotto θ , $W(\theta)$ e $W_e(\theta)$ hanno distribuzione approssimata χ_p^2 , mentre $r(\theta)$ ha distribuzione approssimata $N(0, 1)$.

Tali approssimazioni, sono dette approssimazioni del primo ordine. Possono tuttavia fornire risultati non soddisfacenti dal punto di vista della precisione. Per migliorarla, si può ricorrere alle cosiddette approssimazioni di ordine superiore. In particolare, viene introdotta una nuova quantità (si veda ad esempio Brazzale, Davison e Reid, 2007, paragrafo 2.2), detta radice con segno modificata del log-rapporto di verosimiglianza

$$r^*(\theta) = r(\theta) + \frac{1}{r(\theta)} \log \left\{ \frac{q(\theta)}{r(\theta)} \right\}, \quad (2.8)$$

dove $q(\theta)$ è dato da r_e valutato nella parametrizzazione canonica in una famiglia esponenziale monparametrica. Più in generale, può essere rappresentato da una statistica di tipo *score*.

r^* ha, sotto θ , distribuzione approssimata $N(0, 1)$.

2.2.3 Log-verosimiglianza profilo e risultati asintotici connessi

Spesso, nelle applicazioni, i modelli possiedono svariati parametri, dei quali solo alcuni di diretto interesse per la ricerca. I restanti, necessari affinché il modello sia realistico, assumono tuttavia un'importanza secondaria. Si dirà dunque che il vettore dei parametri $\theta = (\phi, \lambda)$ di dimensione p è costituito da k parametri di interesse (indicati con ϕ) e $p-k$ parametri di disturbo (indicati con λ), dove $1 \leq k < p$. Allo scopo di eliminare λ , un approccio abbastanza generale consiste nel sostituirlo con la sua stima di massima verosimiglianza, ottenuta considerando ϕ fissato. La stima così ricavata viene inserita nella funzione di log-verosimiglianza, che assume il nome di log-verosimiglianza profilo

$$l_p(\phi) = \max_{\lambda} l(\phi, \lambda) = l(\phi, \hat{\lambda}_{\phi}).$$

L'inferenza su ϕ può essere quindi condotta, in via approssimata, in modo analogo al caso con verosimiglianza propria introdotto precedentemente. In particolare, riformulando la (1.5), si ottiene

$$W_p(\phi) = 2(l_p(\hat{\phi}) - l_p(\phi)), \quad (2.9)$$

da cui, se $k = 1$

$$r_p(\phi) = \text{sgn}(\hat{\phi} - \phi) \sqrt{W_p(\phi)}. \quad (2.10)$$

W_p si distribuisce approssimativamente, sotto condizioni di regolarità, come un χ_k^2 , r_p ha invece distribuzione approssimata $N(0, 1)$.

Anche in questo caso, in particolare quando il numero dei parametri di disturbo risulta elevato, tali approssimazioni possono non rivelarsi sufficientemente precise. Nuovamente, è possibile ricorrere alla radice con segno modificata del log-rapporto di verosimiglianza, che in questo caso avrà forma

$$r^*(\phi) = r_p(\phi) + \frac{1}{r_p(\phi)} \log \left\{ \frac{q(\phi)}{r_p(\phi)} \right\}, \quad (2.11)$$

dove, per una famiglia esponenziale,

$$q(\phi) = \sqrt{\frac{|j(\hat{\phi}, \hat{\lambda}_\phi)|}{|j_{\lambda\lambda}(\hat{\phi}, \hat{\lambda}_\phi)|}} (\hat{\phi} - \phi) \sqrt{\frac{|j_{\lambda\lambda}(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_\phi)|}},$$

valutato nella parametrizzazione canonica.

r^* ha, sotto θ , distribuzione approssimata $N(0, 1)$. Contiene un aggiustamento per la non normalità di r_p e per la presenza di parametri di disturbo. Per una trattazione più approfondita si rimanda nuovamente a Brazzale, Davison e Reid (2007, paragrafo 2.3).

2.3 Applicazione al caso in esame

2.3.1 Modello

Lo spazio campionario è costituito da \mathbb{N}^n , dove n rappresenta il numero di replicazioni dell'esperimento. Si assume che $y = (y_1, \dots, y_n)$ sia un campione casuale semplice di numerosità n generato da (Y_1, \dots, Y_n) variabili casuali indipendenti e identicamente distribuite aventi distribuzione $Poi(\theta)$.

Il modello statistico è dunque definito come $\mathcal{F} = \left\{ p(y; \theta) : e^{-\theta \frac{\theta y}{y!}}, \theta \in \Theta \right\}$, dove $\theta = \gamma\mu + \beta$; β e γ possono essere assunti costanti note, o può essere necessario ottenerne una stima riformulando però, in tal caso, il modello per garantire l'identificabilità dei parametri.

Verrà dapprima esposto il caso privo di parametri di disturbo; in particolare β sarà assunto noto e, per semplicità e chiarezza d'esposizione, γ sarà assunto noto e pari ad 1.

Successivamente sarà trattato il caso in cui β e γ non sono noti ed è necessario ottenerne una stima. Contestualmente verrà presentato il modello che garantisce l'identificabilità dei parametri.

2.3.2 β noto

Assumendo β noto, si ha la log-verosimiglianza propria

$$l(\mu) = -n(\mu + \beta) + \sum_{i=1}^n y_i \log(\mu + \beta) - \log \prod_{i=1}^n y_i!. \quad (2.12)$$

Calcolando la derivata prima e ponendola pari a zero si ottiene la stima di massima verosimiglianza di θ , basata sui dati y , considerando per il momento lo spazio parametrico naturale $\mu > -\beta$,

$$l_*(\mu) = -n + \frac{\sum_{i=1}^n y_i}{\mu + \beta} \quad (2.13)$$

$$l_*(\mu) = 0 \Leftrightarrow \mu = \frac{\sum_{i=1}^n y_i}{n} - \beta = \hat{\mu}. \quad (2.14)$$

L'informazione osservata risulta

$$j(\mu) = \frac{\sum_{i=1}^n y_i}{(\mu + \beta)^2}. \quad (2.15)$$

Un intervallo di confidenza per μ con livello $1 - \alpha$ si può ottenere mediante

$$\hat{\mu} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{(\hat{\mu} + \beta)^2}{\sum_{i=1}^n y_i}}, \quad (2.16)$$

oppure

$$\left\{ \mu \in M : l(\mu) \geq l(\hat{\mu}) - \frac{1}{2} \chi_{1,1-\alpha}^2 \right\}, \quad (2.17)$$

ossia

$$\left\{ \mu \in M : n(\hat{\mu} - \mu) + \sum_{i=1}^n y_i \log \frac{\mu + \beta}{\hat{\mu} + \beta} \geq -\frac{1}{2} \chi_{1,1-\alpha}^2 \right\}. \quad (2.18)$$

La (2.18) non ha, in generale, soluzione esplicita: una possibile soluzione consiste nel calcolare numericamente $W(\mu)$ per un insieme di valori del parametro, tracciare una curva che interpoli i punti $(\mu, W(\mu))$, ottenendo così in via approssimata la curva di distribuzione di $W(\mu)$; quindi, stabilito il livello di confidenza $1 - \alpha$, leggere i valori di μ corrispondenti a $\chi_{1,1-\alpha}^2$.

2.3.3 Una spiegazione elementare del problema

Si cercherà ora di fornire, attraverso un esempio, una semplice spiegazione del motivo per cui il vincolo fisico posto su μ possa provocare alcune difficoltà nella stima di intervalli di confidenza per il parametro. In particolare, sarà dapprima presentato il problema così come posto dai fisici, ed in seguito accennata la soluzione proposta da Fraser, Reid e Wong (2004) (qui adottata e discussa in forma estesa nel Capitolo 3).

Si costruiranno intervalli di confidenza per μ basandosi sul risultato (2.16) per semplicità nei calcoli e nella spiegazione, analoghe considerazioni possono essere tuttavia estese a risultati ottenuti utilizzando approssimazioni asintotiche differenti.

Si supponga di aver compiuto un singolo esperimento ($n = 1$), di aver riportato $y = 3$ arrivi e di essere a conoscenza del valore di β , pari, ad esempio, a 2. Si ricorda che γ è assunto noto e pari ad 1.

La stima di massima verosimiglianza per μ risulterebbe $\hat{\mu} = 1$. In assenza del vincolo posto su θ ($\theta \geq \beta$), un intervallo di confidenza di livello $1 - \alpha = 0.95$ per μ , utilizzando l'approssimazione della distribuzione dello stimatore con la distribuzione normale, risulterebbe pari a $[-2.39; 4.39]$. Tuttavia, in presenza del vincolo, tale intervallo si ridurrebbe a $[0; 4.39]$, rendendo apparentemente priva di senso l'asserzione sul livello di confidenza espresso.

In alternativa, in accordo con l'ipotesi formulata, è possibile calcolare un intervallo unilaterale sinistro: mantenendo la stessa approssimazione e lo stesso livello di confidenza, il limite superiore risulterebbe circa pari a 3.85. Posto il vincolo, si avrebbe dunque l'intervallo $[0; 3.85]$.

Inoltre, se β fosse sufficientemente elevato e se si registrasse un numero di arrivi sufficientemente ridotto, si protrebbe incorrere in intervalli di confidenza non contenenti valori ammissibili per μ o contenenti solo il valore $\mu = 0$: il livello di confidenza espresso, in tal caso, risulterebbe ulteriormente incerto ed indefinito.

Come osservato con un certo sarcasmo da Lyons (2008): *“Particle Physicists tend to be unhappy when their years of work result in an empty interval*

for the parameter of interest, and it is little consolation to hear that frequentist statisticians are satisfied with this feature, as it does not lead to undercoverage. When n (il numero di arrvi osservato, ndr) is not quite small enough to result in an empty interval, the upper limit might be very small. This could confuse people into thinking that the experiment was much more sensitive than it really was.”

Ora, è evidente che il problema principale non è tanto lo sconforto dei fisici delle particelle, quanto la difficoltà di interpretazione dei risultati ottenuti. Considerarli completamente inammissibili, vorrebbe dire dover riformulare il modello a posteriori o non utilizzare parte delle informazioni raccolte: entrambe le possibilità non sembrano particolarmente soddisfacenti perché potrebbero essere lette come tecniche ‘manipolative’ (e quindi foriere di distorsioni) dei dati osservati.

Seguendo quanto affermato da Fraser, Reid e Wong (2004), il problema presentato, tuttavia, sorge proprio dalla scelta di costruire intervalli di confidenza per valutare l’effettiva presenza del segnale. La presenza del vincolo può infatti portare, come appena descritto, ad un’interpretazione molto difficoltosa dei risultati ottenuti. L’alternativa proposta da Fraser, Reid e Wong (2004) consiste nell’impiego della funzione di verosimiglianza e della *significance function*, che verrà introdotta nel Capitolo 3, come misura dell’evidenza empirica riguardante il parametro.

Come espresso dallo stesso Mandelkern (2002), infatti: *“It may be most appropriate to, at least in ambiguous cases, give up the notion of characterizing experimental uncertainty with a confidence interval and instead [...] to present the likelihood function for this purpose.”*

2.3.4 Presenza di parametri di disturbo

In presenza di parametri di disturbo, come anticipato, è opportuno riformulare il modello per garantire l’identificabilità dei parametri. In tal modo è possibile ottenere le stime di massima verosimiglianza per γ e β che entre-

ranno a far parte della log-verosimiglianza profilo necessaria per la stima di μ . Si adotta nel seguito il modello proposto da Davison e Sartori (2008).

Il parametro $\theta = (\mu, \lambda)$ è vettoriale, μ rappresenta il parametro di interesse dell'esperimento, mentre $\lambda = (\gamma, \beta)$ costituisce il vettore dei parametri di disturbo. Mentre μ è considerato fisso nei diversi canali dell'esperimento, λ assume valori specifici per ciascun canale. Si assume dunque che i dati osservati siano stati generati da tre distribuzioni di Poisson indipendenti Y_{1k}, Y_{2k}, Y_{3k} , di media, rispettivamente, $\gamma_k\mu + \beta_k$, $\beta_k t_k$, $\gamma_k u_k$, $k = 1, \dots, n$. t_k e u_k sono considerate costanti note positive.

Si consideri la riparametrizzazione $\lambda_k = (\lambda_{1k}, \lambda_{2k})$ delle componenti di disturbo (β_k, γ_k) , $k = 1, \dots, n$, $\lambda_{1k} = \log(\beta_k) - \log(\gamma_k)$ e $\lambda_{2k} = \log(\beta_k)$.

Dovendo essere $\gamma_k\mu + \beta_k > 0$ per ogni k , si pone $\mu > \max_k \{-\beta_k/\gamma_k\}$, nella nuova parametrizzazione $\mu > \max_k \{-e^{\lambda_{1k}}\}$, mentre $\lambda_{2k} \in \mathbb{R}$.

La funzione di log-verosimiglianza (somma delle funzioni di verosimiglianza delle tre osservazioni e degli n canali) assume la forma

$$l(\theta) = \sum_{k=1}^n \{a_k(\theta)^T y_k - c_k(\theta)\}, \quad (2.19)$$

dove

$$\begin{aligned} a_k(\theta)^T &= \{\log(\mu e^{\lambda_{2k} - \lambda_{1k}} + e^{\lambda_{2k}}), \lambda_{2k}, (\lambda_{2k} - \lambda_{1k})\} \\ y_k &= (y_{1k}, y_{2k}, y_{3k})^T \\ c_k(\theta) &= (\mu + u_k) e^{\lambda_{2k} - \lambda_{1k}} + (1 + t_k) e^{\lambda_{2k}}. \end{aligned}$$

Per ottenere la funzione di verosimiglianza profilo per μ , si otterrà dapprima la stima di massima verosimiglianza per i parametri di disturbo fissato μ . Successivamente, la stima di massima verosimiglianza per μ si ricaverà tramite la massimizzazione della log-verosimiglianza profilo $l_p(\mu)$. In generale, tali valori non saranno ottenibili in forma esplicita, ma dovranno essere calcolati numericamente. Gli intervalli di confidenza potranno quindi essere ricavati mediante le diverse approssimazioni disponibili: nel Capitolo 3 verranno in

particolare presentati intervalli ottenuti mediante $r(\theta)$, ma un'inferenza più precisa può basarsi su $r^*(\theta)$ (Davison e Sartori, 2008).

Anche in questo caso, benché la formulazione del modello sia più complessa, in presenza di un numero ridotto di arrivi è possibile incorrere in intervalli di confidenza parzialmente o completamente esterni all'intervallo dei valori ammissibili per il parametro: l'unico vincolo posto è, infatti, $\gamma_k\mu + \beta_k > 0$.

Nuovamente, la funzione di verosimiglianza e la *significance function* forniscono una soluzione al problema proposto.

Si cercherà ora di introdurre teoricamente ed operativamente tale soluzione.

Capitolo 3

Significance functions

“So che non so quel che non so; invidio coloro che sapranno di più, ma so che anch’essi, come me, avranno da misurare, pesare, dedurre e diffidare delle deduzioni ottenute, stabilire nell’errore qual è la parte del vero e tener conto nel vero dell’eterna presenza di falso.

[...] Mi sono guardato bene dal fare della verità un idolo; ho preferito lasciarle il nome più umile di esattezza.”

M. Yourcenar, L’opera al nero

3.1 Un diverso modo di pensare

Si supponga di voler ottenere una misura statistica dell’evidenza empirica riguardante un parametro θ .

Lo scopo è generalmente di ottenere una stima puntuale del parametro θ incognito, ed eventualmente intervalli di confidenza e test.

La stima puntuale può essere ottenuta attraverso diversi metodi, tra i quali assume qui particolare rilevanza la massimizzazione della funzione di verosimiglianza, secondo le modalità descritte nel Capitolo 2. Nel caso in cui tale stima venga interpretata come funzione dei dati osservati e non sempli-

come valore numerico, rappresenta a sua volta una variabile casuale dotata di una propria distribuzione di probabilità, e viene detta stimatore.

Qualora sia nota la distribuzione esatta dello stimatore, è possibile calcolare agevolmente intervalli di confidenza e test per θ : la distribuzione dello stimatore dipenderà infatti a sua volta da θ , e la stima ottenuta può essere vista come una realizzazione, appunto, della variabile casuale stimatore.

Purtroppo, la distribuzione esatta non è generalmente nota, ad esclusione di alcuni semplici casi. Per questo, si ricorre alle approssimazioni introdotte nel Capitolo 2, in particolare approssimazioni del primo ordine e, qualora risulti necessario per ottenere una maggiore precisione, approssimazioni di ordine superiore. Il maggior grado di precisione di queste ultime è stato verificato anche empiricamente confrontando le distribuzioni approssimate così ottenute con le distribuzioni esatte, dove disponibili (si veda Brazzale, Davison e Reid, 2007).

Attuando una verifica di ipotesi, assume un ruolo centrale il p -value osservato, che, si ricorda, corrisponde alla “probabilità di ottenere un risultato pari o più estremo di quello osservato, sotto l’ipotesi nulla”; quindi, volendo verificare un determinato valore di θ , sia $\bar{\theta}$, questo viene ipotizzato il vero parametro della distribuzione che ha generato i dati, e attraverso il p -value si ottiene una misura della probabilità con cui si otterrebbe il risultato sperimentale osservato, o uno più estremo, se $\bar{\theta}$ fosse, effettivamente, il vero valore del parametro incognito.

In una singola verifica di ipotesi è calcolato un unico p -value, connesso al valore oggetto di verifica. Tuttavia, è possibile estendere il medesimo procedimento a tutti i valori di θ appartenenti allo spazio parametrico: in tal modo, si otterrà un p -value per ogni possibile valore di θ . Secondo questa logica è costruita la *significance function*. Essa può essere applicata a distribuzioni esatte ed approssimate, e fornisce una misura statistica complessiva dell’informazione disponibile riguardante il parametro.

Per un riferimento più dettagliato al procedimento descritto si veda Fraser (1992) e Fraser, Reid e Wong (2004).

3.1.1 Un esempio: la media di una distribuzione normale

Verrà ora presentato un semplice esempio allo scopo di comprendere come si possa costruire, operativamente, una *significance function* (Fraser, Reid e Wong, 2004).

Si assume che $y = (y_1, \dots, y_n)$ sia un campione casuale semplice di numerosità n generato da (Y_1, \dots, Y_n) variabili casuali indipendenti e identicamente distribuite aventi distribuzione $N(\mu, \sigma_0^2)$, con σ_0^2 assunto noto. Sia \bar{y}_n la media campionaria.

La verosimiglianza propria, funzione del solo parametro μ , assume la forma

$$L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma_0^2} \right\}. \quad (3.1)$$

La stima di massima verosimiglianza per μ risulta pari a \bar{y}_n . È dunque possibile determinare la distribuzione esatta dello stimatore, che risulta $N(\mu, \sigma_0^2/n)$.

La *significance function* si ottiene calcolando, al variare di μ , la probabilità di osservare un risultato campionario pari o inferiore a \bar{y}_n ; essendo la distribuzione di \bar{y}_n normale, la *significance function* è così definita:

$$p(\mu) = \Phi \left(\frac{\sqrt{n}(\bar{y}_n - \mu)}{\sigma_0} \right), \quad (3.2)$$

dove Φ rappresenta funzione di ripartizione di una normale standard. Log-verosimiglianza normalizzata e *significance function* per $\bar{y}_n = 5$ e $\sigma_0^2 = 2$, si possono osservare nelle Figure 3.1 e 3.2.

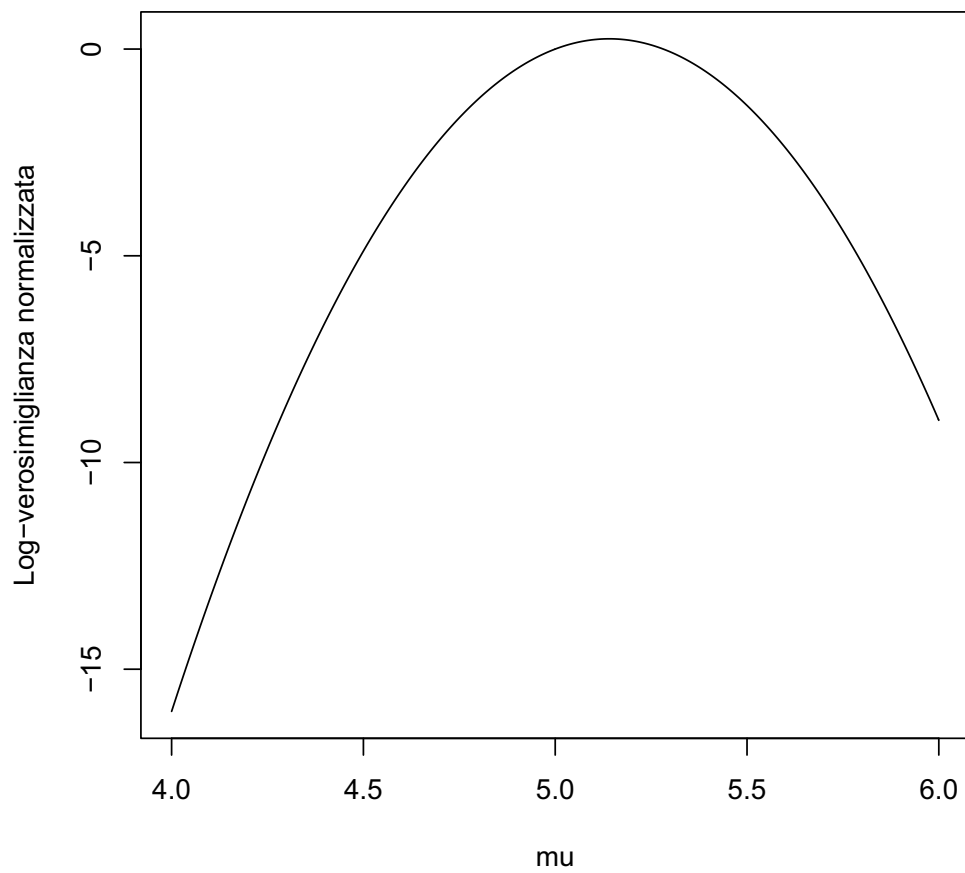


Figura 3.1: *Funzione di log-verosimiglianza normalizzata per μ . $\bar{y}_n = 5$, $\sigma_0^2 = 2$ e $n = 100$.*

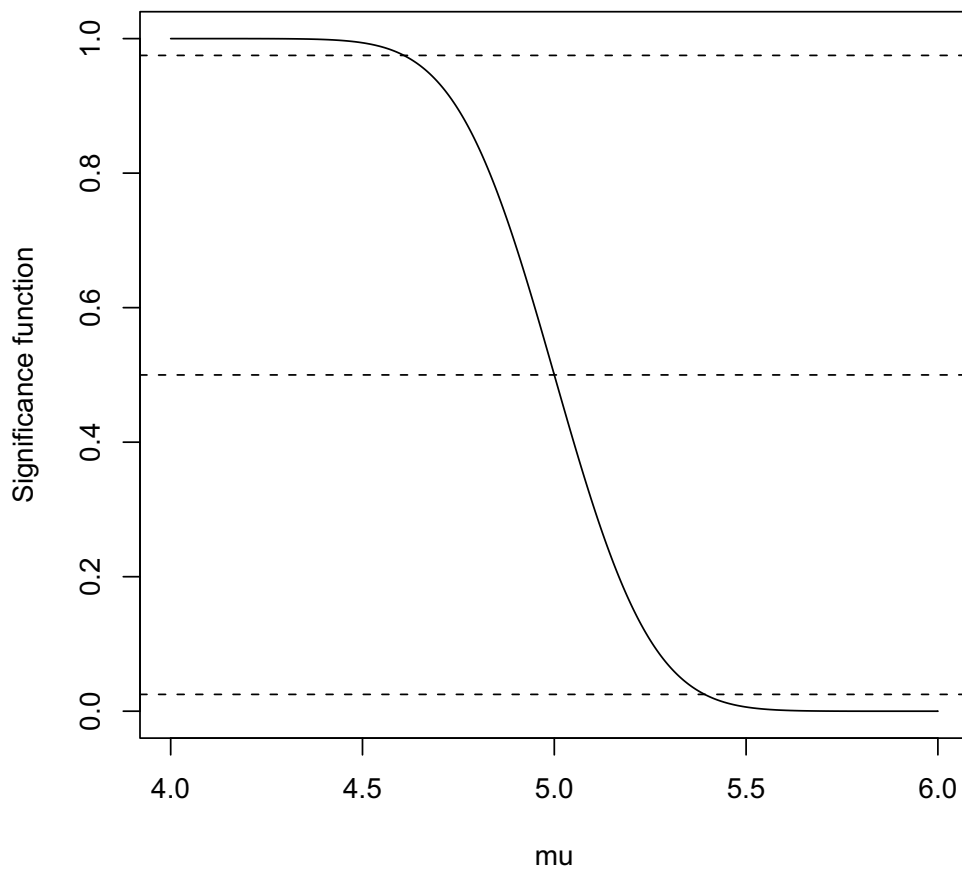


Figura 3.2: *Significance function per μ . $\bar{y}_n = 5$, $\sigma_0^2 = 2$ e $n = 100$. Le linee orizzontali sono in corrispondenza dei valori 0.025, 0.975 e 0.5, fornendo, rispettivamente, i limiti per un intervallo di confidenza con livello $1-\alpha = 0.95$ e una stima non distorta in mediana di μ .*

Come si può osservare dai grafici riportati, la *significance function* permette di ottenere intervalli di confidenza e p -value in modo relativamente semplice e diretto: i limiti di un intervallo di confidenza di livello $1 - \alpha$ sono infatti ottenibili calcolando i valori $\mu_{\alpha/2}$ e $\mu_{1-\alpha/2}$ soluzioni delle equazioni $p(\mu_{\alpha/2}) = 1 - \alpha/2$ e $p(\mu_{1-\alpha/2}) = \alpha/2$.

Alternativamente, il p -value connesso alla verifica di un valore di μ , sia $\bar{\mu}$, con ipotesi alternativa unilaterale destra, è ricavabile dal complemento ad uno del valore che la funzione assume in $\bar{\mu}$, ossia mediante $1 - p(\bar{\mu})$. p -value nel caso di ipotesi alternativa unilaterale sinistra e bilaterale sono analogamente ottenibili senza complicazioni di rilievo.

Inoltre, il valore di μ in cui la funzione assume valore 0.5, rappresenta una stima non distorta in mediana del parametro di interesse.

Si può dunque comprendere perché la *significance function* fornisca una misura statistica complessiva dell'informazione disponibile riguardante il parametro.

3.2 Applicazione al caso in esame

Come introdotto nel Capitolo 2, il problema è di individuare l'evidenza relativa a μ , parametro incognito di una $Poi(\gamma\mu + \beta)$. Anche in questo caso sarà opportuno scindere le due possibili specificazioni del problema, che dipendono dalla presenza o meno di parametri di disturbo.

3.2.1 β, γ noti

Qualora μ rappresenti l'unico parametro ignoto della distribuzione, la formulazione della corrispondente *significance function* può basarsi sulla distribuzione esatta dello stimatore.

Si dovrà tuttavia tenere conto della natura discreta della variabile: per il calcolo dei p -value sarà dunque opportuno utilizzare alcuni accorgimenti.

La verosimiglianza propria assume la forma

$$L(\mu) = \prod_{i=1}^n e^{-(\gamma\mu+\beta)} \frac{(\gamma\mu+\beta)^{y_i}}{y_i!}. \quad (3.3)$$

La stima di massima verosimiglianza per μ risulta pari a

$$\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n\gamma} - \frac{\beta}{\gamma}. \quad (3.4)$$

La *significance function* per μ , come anticipato, può essere ottenuta attraverso la distribuzione esatta dello stimatore.

Posto, per semplicità di notazione, $\theta = \gamma\mu + \beta$ (quindi $\hat{\theta} = \sum_{i=1}^n y_i/n$), si ha infatti $Y_i \sim Poi(\theta)$, da cui segue $\sum_{i=1}^n Y_i \sim Poi(n\theta)$. Sia y_0 , inoltre, il numero totale di arrivi osservato nelle n replicazioni dell'esperimento.

La natura discreta della variabile implica che, volendo attuare un'approssimazione ad una distribuzione continua, il semplice calcolo della probabilità cumulata sino ad y_0 come misura del p -value non è appropriato: è infatti necessario tenere conto del fatto che la distribuzione di probabilità discreta cumula in specifici valori una misura di probabilità altrimenti rappresentata da un'area. Per ovviare a tale problema, è possibile utilizzare due espedienti:

1. Il calcolo di un intervallo di p -value;
2. Il calcolo di un *mid p-value*.

Utilizzando il primo metodo, l'intervallo avrà estremi (Fraser, Reid e Wong, 2004)

$$Pr_{\theta} \left(\sum_{i=1}^n Y_i < y_0 \right) = \sum_{z=0}^{y_0-1} e^{-n\theta} \frac{(n\theta)^z}{z!}, \quad (3.5)$$

e

$$Pr_{\theta} \left(\sum_{i=1}^n Y_i \leq y_0 \right) = \sum_{z=0}^{y_0} e^{-n\theta} \frac{(n\theta)^z}{z!}. \quad (3.6)$$

La *significance function* avrà dunque forma

$$p(\theta) = [F^-(y_0, \theta), F(y_0, \theta)], \quad (3.7)$$

dove $F^-(y_0, \theta)$ ed $F(y_0, \theta)$ corrispondono, rispettivamente, ai risultati 3.5 e 3.6.

In alternativa, è possibile ottenere un *mid p-value* (Brazzale, Davison e Reid, 2007, paragrafo 3.3) mediante

$$Pr_\theta \left(\sum_{i=1}^n Y_i < y_0 \right) + \frac{1}{2} Pr_\theta \left(\sum_{i=1}^n Y_i = y_0 \right) = \sum_{z=0}^{y_0-1} e^{-n\theta} \frac{(n\theta)^z}{z!} + \frac{1}{2} e^{-n\theta} \frac{(n\theta)^{y_0}}{y_0!}, \quad (3.8)$$

da cui segue la *significance function*

$$p(\theta) = F^-(y_0, \theta) + \frac{1}{2} f(y_0, \theta). \quad (3.9)$$

Pur essendo disponibile la distribuzione esatta, può essere interessante comprendere come si comporterebbero, in questo semplice caso, le diverse distribuzioni approssimate disponibili.

In primo luogo, è opportuno notare che, per il teorema del limite centrale, per θ sufficientemente grande, si ha $\sum_{i=1}^n Y_i \sim N(n\theta, n\theta)$, da cui segue (Fraser, Reid e Wong, 2004)

$$\sqrt{\sum_{i=1}^n Y_i} \sim N(\sqrt{n\theta}, 1/4), \quad (3.10)$$

che ha il pregio di stabilizzare la varianza. Alternativamente, è possibile utilizzare (Pace e Salvan, 1996, paragrafo 8.5)

$$\left(\sum_{i=1}^n Y_i \right)^{2/3} \sim N((n\theta)^{2/3}, 4/9(n\theta)^{1/3}), \quad (3.11)$$

che migliora, invece, la simmetria.

Anche in questo caso, data la natura discreta della variabile, è opportuno ricavare un intervallo di *p-value*: si utilizzerà la correzione di continuità, le approssimazioni saranno cioè valutate in $y_0 \pm 0.5$.

Le *significance function* assumono in questo caso la forma

$$p(\theta) = \left[\Phi \left(2 \left(\sqrt{y_0 - 0.5} - \sqrt{n\theta} \right) \right), \Phi \left(2 \left(\sqrt{y_0 + 0.5} - \sqrt{n\theta} \right) \right) \right], \quad (3.12)$$

per la trasformata in 3.10, e

$$p(\theta) = \left[\Phi \left(\frac{3/2 \left((y_0 - 0.5)^{2/3} - (n\theta)^{2/3} \right)}{(n\theta)^{1/6}} \right), \Phi \left(\frac{3/2 \left((y_0 + 0.5)^{2/3} - (n\theta)^{2/3} \right)}{(n\theta)^{1/6}} \right) \right], \quad (3.13)$$

per la trasformata in 3.11.

In alternativa, si possono utilizzare le statistiche r_e , r , r^* introdotte nel Capitolo 2. Il computo di $r_e(\theta)$ necessario alla stima di $r^*(\theta)$ avviene però, in questo caso, utilizzando la parametrizzazione canonica delle famiglie esponenziali, in particolare, si avrà $\log(\theta) = \psi$, da cui

$$l^\Psi(\psi) = y_0\psi - e^\psi. \quad (3.14)$$

A partire dalla funzione di verosimiglianza, si possono ricavare la stima di massima verosimiglianza e l'informazione osservata, che risultano, rispettivamente

$$\hat{\psi} = \log \hat{\theta} \quad (3.15)$$

$$j^\Psi(\hat{\psi}) = \exp^{\hat{\psi}} = \hat{\theta}. \quad (3.16)$$

r_e risulta, dunque

$$r_e(\theta) = \left(\log \frac{\hat{\theta}}{\theta} \sqrt{\hat{\theta}} \right). \quad (3.17)$$

Si possono quindi ottenere

$$r(\theta) = \text{sgn}(\hat{\theta} - \theta) \sqrt{2 \left(n(\theta - \hat{\theta}) + y_0 \log \frac{\hat{\theta}}{\theta} \right)}, \quad (3.18)$$

e

$$r^*(\theta) = r(\theta) + \frac{1}{r(\theta)} \log \left\{ \frac{r_e(\theta)}{r(\theta)} \right\}. \quad (3.19)$$

Ricordando che esse assumono, sotto θ , distribuzione approssimata $N(0, 1)$, le *significance function* corrispondenti sono rappresentate da $\Phi(r_e)$, $\Phi(r)$ e $\Phi(r^*)$.

Riprendendo l'esempio presentato nel paragrafo 2.3.3, si supponga di avere compiuto $n = 1$ esperimenti, di aver riportato $y_0 = 3$ arrivi e di essere a conoscenza del valore di $\beta = 2$. Inoltre, per semplicità, si supponga γ noto e pari ad 1.

In Figura 3.3 si riporta la verosimiglianza per μ . In Figura 3.4 si confrontano le *significance function* ottenute attraverso il *mid p-value* della distribuzione esatta e le due approssimazioni alla distribuzione normale (valutate in $y_0 \pm 0.5$). Infine, in Figura 3.5 si confrontano le *significance function* date dalla distribuzione esatta (*mid p-value*), da r ed r^* .

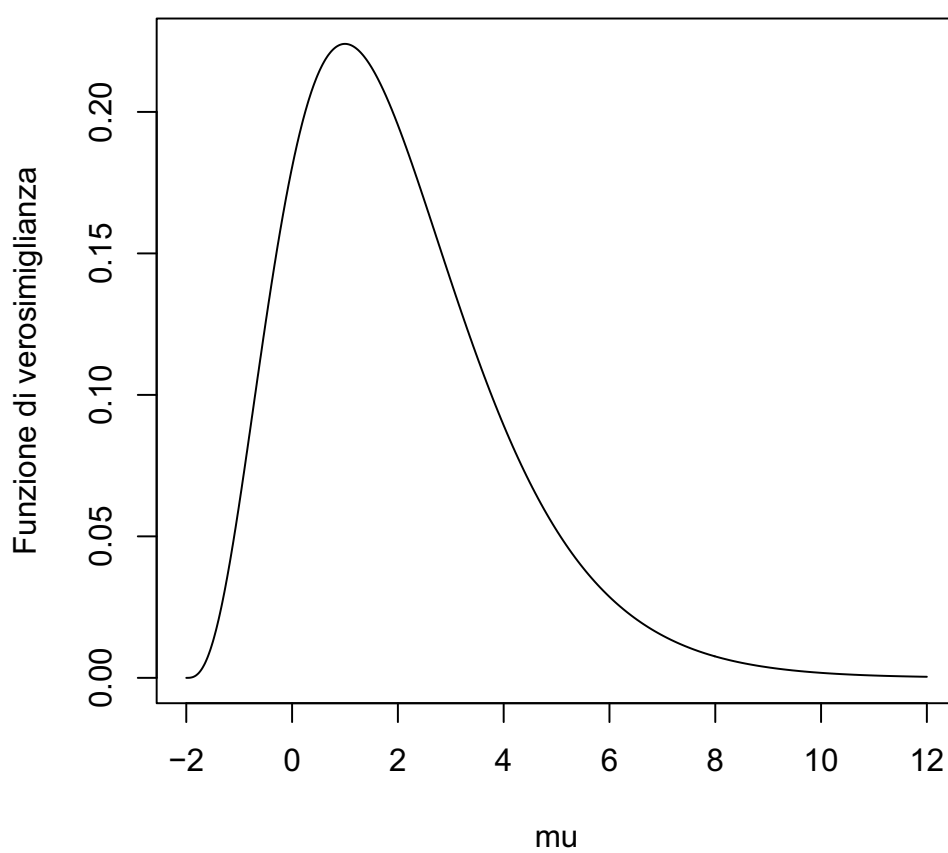


Figura 3.3: *Funzione di verosimiglianza per μ . $y_0 = 3$, $\beta = 2$ e $\gamma = 1$.*

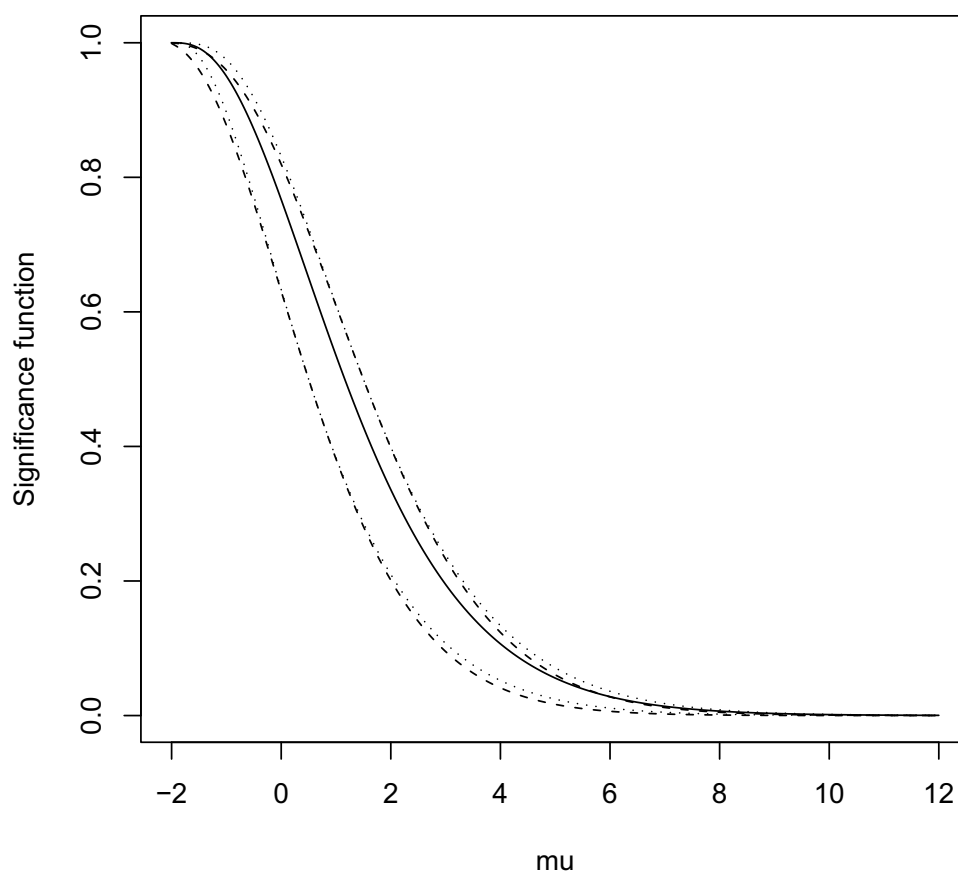


Figura 3.4: *Significance function per μ : distribuzione esatta - mid p-value (linea continua), approssimazione normale con $\sqrt{y_0 \pm 0.5}$ (linea tratteggiata), approssimazione normale con $(y_0 \pm 0.5)^{2/3}$ (punti). $y_0 = 3$, $\beta = 2$ e $\gamma = 1$.*

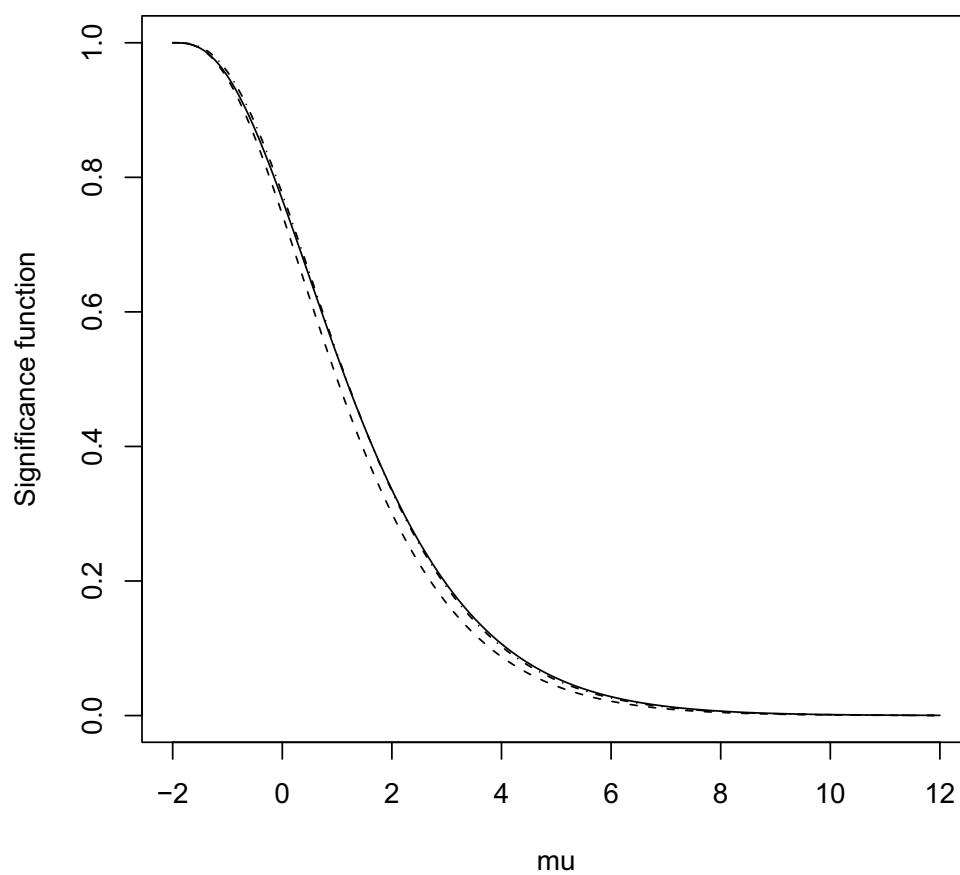


Figura 3.5: *Significance function per μ : distribuzione esatta - mid p-value (linea continua), r (linea tratteggiata), r^* (punti). $y_0 = 3$, $\beta = 2$ e $\gamma = 1$.*

Come si può osservare dai grafici, le approssimazioni alla distribuzione normale tendono a fornire un intervallo di p -value piuttosto ampio ed impreciso rispetto ai p -value forniti dalla distribuzione esatta, soprattutto nel suo limite inferiore.

Nel secondo grafico si può osservare il comportamento delle distribuzioni approssimate, in particolare, la precisione raggiunta da r^* , che risulta quasi perfettamente sovrapposto alla distribuzione esatta: l'errore è stato infatti portato dall'ordine di $\mathcal{O}(n^{-1/2})$ (approssimazioni basate sul teorema del limite centrale), all'ordine di $\mathcal{O}(n^{-3/2})$ utilizzando r^* .

Le approssimazioni, soprattutto quelle legate alla convergenza in distribuzione alla distribuzione normale, risultano ulteriormente precise all'aumentare del numero di eventi osservati: per comprendere anche visivamente tale fenomeno, si supponga di compiere un nuovo esperimento, nel quale vengano osservati $y_0 = 30$ arrivi, β è supposta pari a 6, $\gamma = 3$. I grafici corrispondenti sono riportati nelle Figure 3.6, 3.7 ed 3.8.

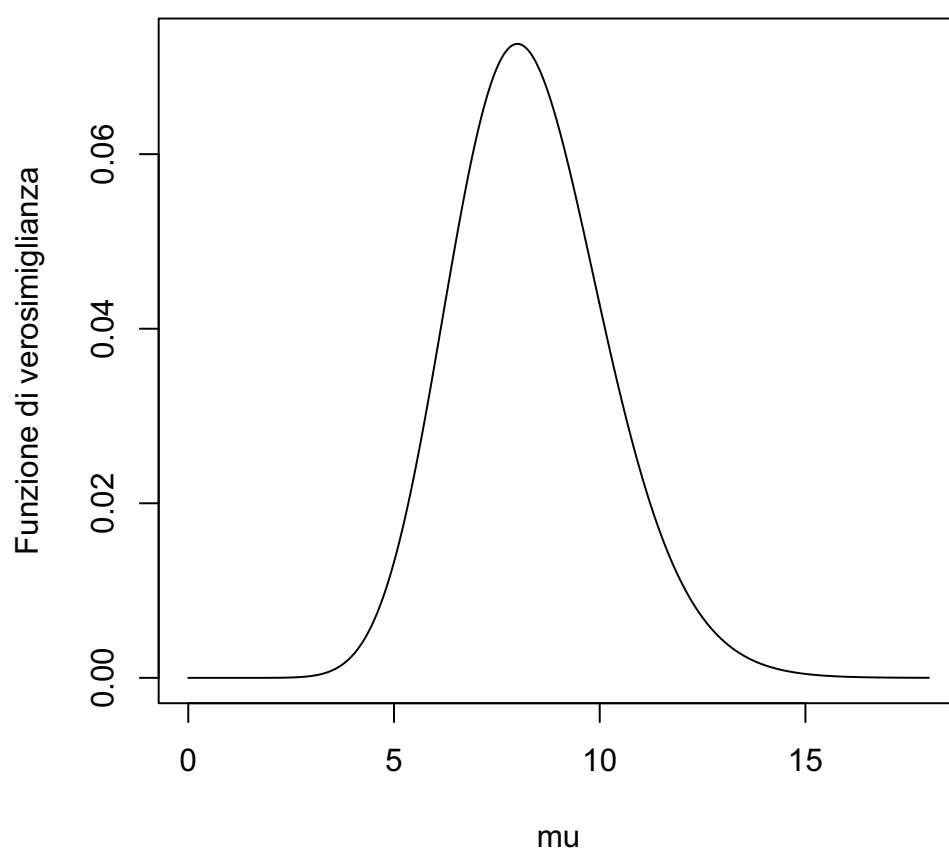


Figura 3.6: *Funzione di verosimiglianza per μ . $y_0 = 30$, $\beta = 6$ e $\gamma = 3$.*

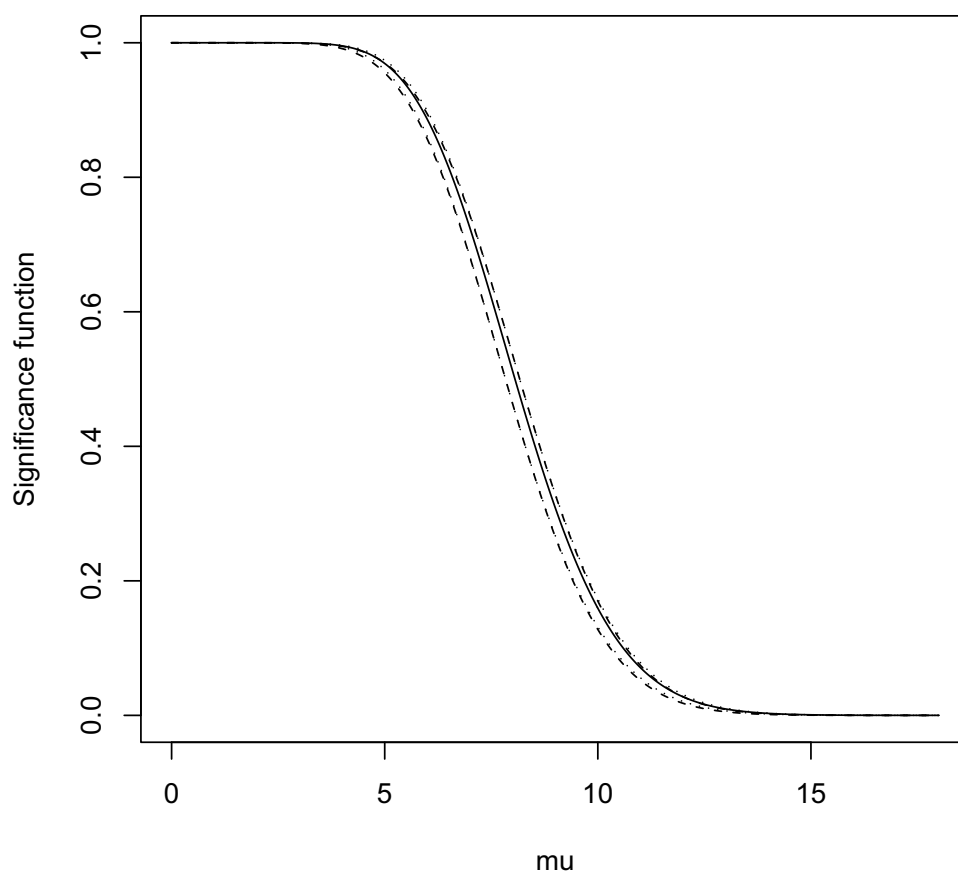


Figura 3.7: *Significance function per μ : distribuzione esatta - mid p-value (linea continua), approssimazione normale con $\sqrt{y_0 \pm 0.5}$ (linea tratteggiata), approssimazione normale con $(y_0 \pm 0.5)^{2/3}$ (punti). $y_0 = 30$, $\beta = 6$ e $\gamma = 3$.*

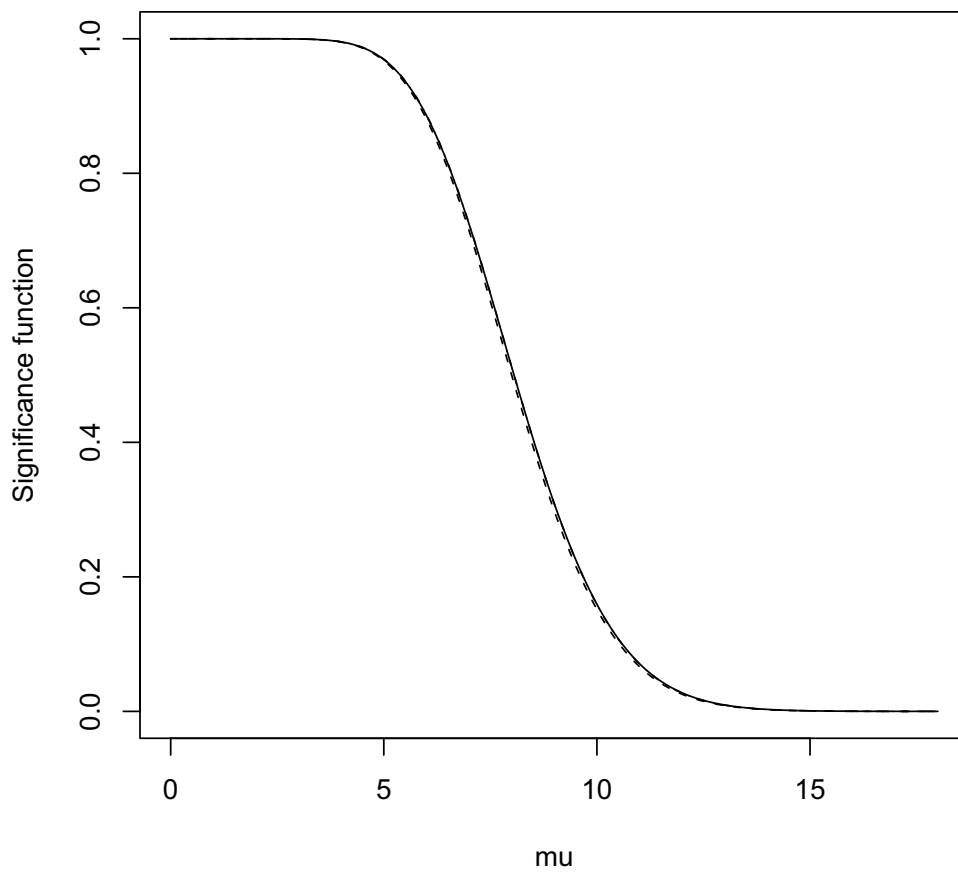


Figura 3.8: *Significance function per μ : distribuzione esatta - mid p-value (linea continua), r (linea tratteggiata), r^* (punti). $y_0 = 30$, $\beta = 6$ e $\gamma = 3$.*

Aumentando la numerosità, diminuisce inoltre, fino ad annullarsi, la probabilità di incorrere in intervalli di confidenza per μ parzialmente esterni alla regione dei valori ammissibili. Per questo motivo, quando il numero di arrivi è relativamente elevato, il problema risulta di scarsa complessità.

È utile a questo punto comprendere in cosa consista la soluzione fornita dalla *significance function*.

La *significance function* fornisce p -value per ogni possibile valore del parametro, nel caso specifico per $\mu \in [-\beta/\gamma, \infty)$. Nel Capitolo 2 è stato tuttavia sottolineato il problema introdotto dal vincolo posto su μ , che deve risultare, ragionevolmente, maggiore o uguale a 0.

Il vantaggio della soluzione proposta deriva dal fatto che la determinazione di un p -value per valori di μ ‘plausibili’, appartenenti cioè alla regione $[0, \infty)$, non presenta ambiguità: verificando, ad esempio, l’ipotesi $\mu = 0$ contro $\mu > 0$, il p -value corrispondente si otterrà semplicemente tramite $1 - p(\theta) |_{\theta=\beta}$.

I p -value ottenuti, invece, per valori appartenenti all’intervallo $[-\beta/\gamma, 0)$, possono essere utilizzati, seguendo la proposta di Fraser, Reid e Wong (2004), come misura diagnostica relativa al valore assunto di β e γ , e relativa al modello (suggerendone l’eventuale erroneità).

3.2.2 Presenza di parametri di disturbo

Qualora γ e β siano ignoti, il modello deve essere riformulato come descritto nel paragrafo 2.3.4. In questo caso, assumendo lo stimatore una forma più complessa, i p -value possono essere calcolati solo mediante la sua distribuzione approssimata.

Inoltre, come anticipato, non è possibile ottenere una formula esplicita per la stima dei parametri di disturbo e di μ . Ci si limiterà dunque ad illustrare i passaggi ed i risultati ottenuti attraverso un esempio numerico.

I dati utilizzati non sono, purtroppo, dati reali: come accennato nel Capitolo 1, l’LHC ha solo recentemente ripreso a funzionare, e la raccolta dei dati è al momento solo agli inizi.

Si utilizzeranno quindi i dati proposti nel corso del Banff Workshop, che erano stati simulati senza rendere noti i veri valori dei parametri ai partecipanti: una volta scelta la metodologia ed effettuate le analisi, era verificato il livello di copertura degli intervalli ottenuti.

Il data-set è costituito da 60229 record (ciascuno corrispondente ad un campione), e cinque colonne, rappresentanti, rispettivamente, y_{1k} (il numero di arrivi della misurazione principale), y_{2k} (il numero di arrivi della misurazione necessaria per la stima del background), y_{3k} (il numero di arrivi della misurazione necessaria per la stima dell'efficienza dello strumento), t_k ed u_k (costanti note). Nel caso proposto, inoltre si ha un unico canale, per cui $n = 1$.

La stima di massima verosimiglianza per μ risulta pari a circa 13.093, mentre β e γ sono stimate rispettivamente pari a 3.013 e 0.993.

La funzione di log-verosimiglianza normalizzata per μ e la significance function calcolata mediante r_p , sono riportate nelle Figure 3.9 e 3.10.

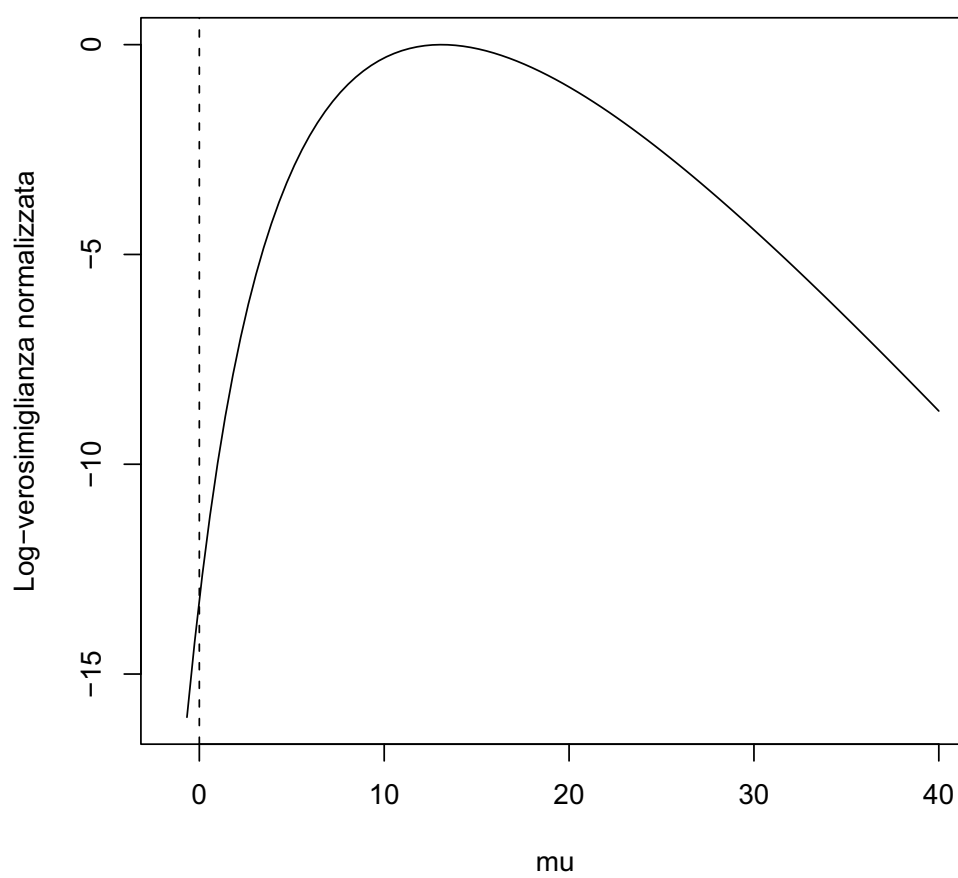


Figura 3.9: *Log-verosimiglianza normalizzata per μ : $\beta = 3.013$, $\gamma = 0.993$.*

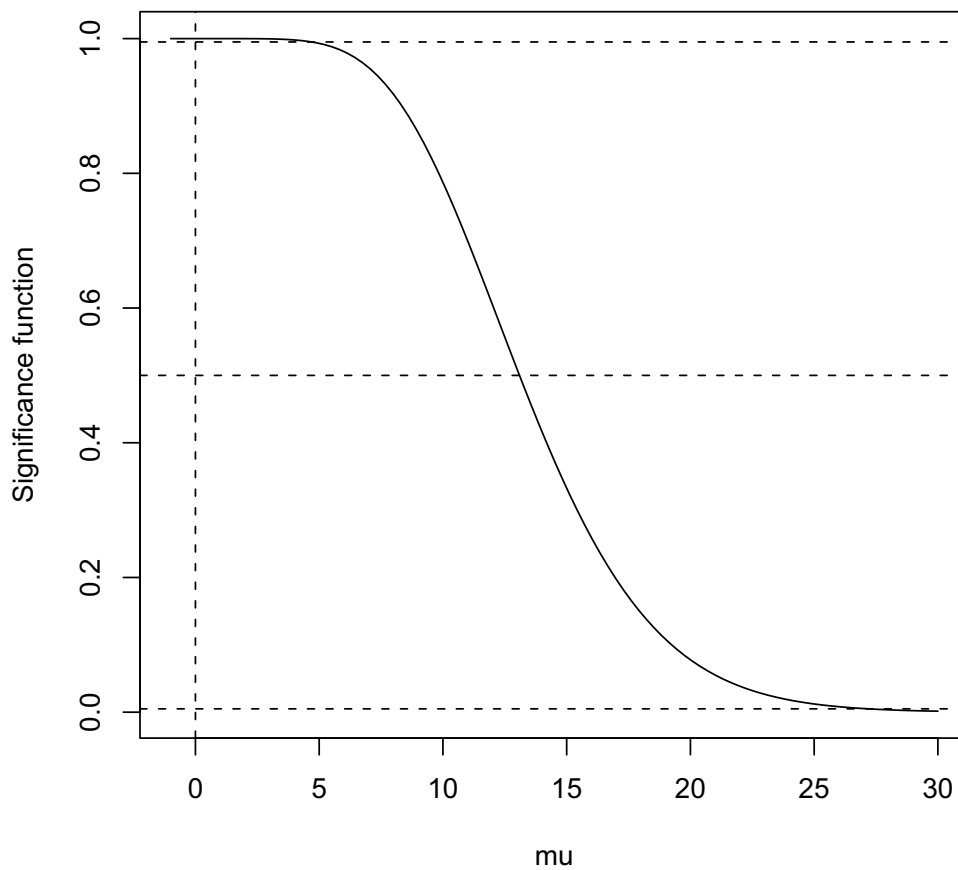


Figura 3.10: *Significance function per μ : $\beta = 3.013$, $\gamma = 0.993$. Le linee orizzontali sono in corrispondenza dei valori 0.005 e 0.995 (individuando così i limiti di un intervallo di confidenza con livello $1 - \alpha = 0.99$), e 0.5 (corrispondente ad una stima non distorta in mediana di μ). La linea verticale individua il p-value per verificare l'ipotesi $\mu = 0$.*

Ai partecipanti era richiesto il calcolo di intervalli di confidenza e limiti superiori per μ allo scopo di verificare successivamente, come accennato, il livello di copertura raggiunto. Questi possono essere agevolmente ottenuti tramite la *significance function*: fissato un livello $1 - \alpha$ di copertura teorica, si possono ottenere intervalli bilaterali $(\mu_{\alpha/2}, \mu_{1-\alpha/2})$ risolvendo le equazioni $\Phi(r_p(\mu_{\alpha/2})) = 1 - \alpha/2$ per il limite superiore, e $\Phi(r_p(\mu_{1-\alpha/2})) = \alpha/2$ per il limite inferiore.

È ancora possibile tuttavia incorrere in limiti che non rispettano il vincolo fisico $\mu > 0$: seguendo Davison e Sartori (2008), se il limite inferiore, ad esempio, è negativo, si prenderà come limite il massimo $\max(\mu_{1-\alpha/2}, 0)$.

Allo stesso modo, è anche possibile ottenere solo limiti inferiori o superiori, in accordo con l'ipotesi unilaterale. Questi avranno forma, rispettivamente: $[\mu_{1-\alpha}, \infty]$ e $[0, \mu_\alpha]$, con $\mu_{1-\alpha}$ e μ_α soluzioni, rispettivamente, delle equazioni $\Phi(r_p(\mu_\alpha)) = \alpha$ e $\Phi(r_p(\mu_\alpha)) = 1 - \alpha$.

Benché l'interesse si sia focalizzato nel corso dell'incontro soprattutto sul calcolo di limiti superiori, è sottolineato (Davison e Sartori, 2008) come sia in realtà più idoneo per verificare la presenza o meno di segnali di interesse, il calcolo di un limite inferiore.

Infine, è possibile non ricorrere del tutto ad intervalli di confidenza, e basare le proprie considerazioni solo sul p -value ottenuto: verificando l'ipotesi $\mu = 0$ contro $\mu > 0$, esso può essere facilmente ricavato, nel caso qui presentato, tramite $1 - \Phi(r_p(0))$, e risulta pari a circa 1.332×10^{-7} , suggerendo il rifiuto dell'ipotesi nulla ad un livello lievemente inferiore al canonico - nell'ambito degli esperimenti fisici - $1 - \Phi(5)$ presentato nel Capitolo 1.

Nel caso in cui, inoltre, si ottenessero valori negativi come limiti dell'intervallo, il p -value sarebbe molto prossimo ad 1, validando così l'ipotesi di assenza di segnali di interesse. Ciononostante, per quanto potrebbero comunque essere tratte conclusioni sull'esito dell'esperimento, sarebbe in tal caso opportuno interrogarsi sulla bontà del modello utilizzato.

Conclusioni

Una possibile risposta al problema fisico e, conseguentemente, statistico proposto sembra essere dunque costituita dalla *significance function* presentata nel Capitolo 3. Essa fornisce infatti una misura complessiva dell'informazione disponibile riguardante il parametro, dati gli eventi osservati ed il modello scelto.

La sua applicazione alle differenti statistiche disponibili fornisce risultati dotati di un diverso grado di precisione: nel caso in cui è stato possibile confrontare la *significance function* ricavata dalla distribuzione esatta (caso privo di parametri di disturbo), è stato osservato, infatti, che le distribuzioni approssimate basate sul teorema del limite centrale tendono a rivelarsi alquanto imprecise, soprattutto quando il numero di eventi osservato risulta particolarmente ridotto. La radice con segno e la radice con segno modificata del log-rapporto di verosimiglianza, invece, tendono ad approssimare con un grado di precisione più elevato la distribuzione esatta. In presenza di un numero più elevato di arrivi, tutte le approssimazioni presentano un migliore comportamento.

In presenza di parametri di disturbo, il modello è stato riformulato e non è stato possibile individuare la distribuzione esatta dello stimatore, né fornire stimatori in forma esplicita per i parametri di disturbo e di interesse. A scopo esemplificativo, è stata calcolata la *significance function* relativa alla statistica r_p , comunque ottenibile in modo diretto e relativamente semplice, una volta ricavata la verosimiglianza profilo per il parametro di interesse ed una sua stima numerica di massima verosimiglianza.

È stato osservato in Davison e Sartori (2008) che stime puntuali per il parametro di interesse leggermente migliori possono essere ottenute attraverso un approccio bayesiano con a priori non informativa. La trattazione esula però dagli scopi di questa tesi.

In conclusione, benché sia tecnicamente possibile incorrere in limiti negativi per gli intervalli di confidenza relativi al parametro di interesse, la *significance function* permette di calcolare p -value privi di ambiguità per ogni valore posto sotto verifica. Ciononostante, valori del parametro altamente plausibili, ma esterni alla regione ragionevolmente (fisicamente) ammissibile, dovrebbero indurre a riconsiderare la validità del modello ipotizzato e la stima (o il valore assunto) dei restanti parametri.

Bibliografia

- ANDRADE, E. N. DA C. (1964). *Rutherford and the Nature of the Atom*, Anchor Books, New York.
- BERGER, J. (2008). A Comparison of testing methodologies. In PHYSTAT-LHC Workshop at CERN on Statistical Issues for LHC Physics. Disponibile a: <http://phystat-lhc.web.cern.ch/phystat-lhc/2008-001.pdf>.
- BOLTZMANN, L. (2010). trad. e cura di BADINO, M., *Fisica e Probabilità*, Melquìades edizioni, Milano.
- BRAZZALE, A. R., DAVISON, A. C. e REID, N. (2007). *Applied Asymptotics: Case Studies in Small Sample Statistics*. Cambridge Univ. Press, Cambridge.
- DAVISON, A. C. e SARTORI, N. (2008). The Banff Challenge: statistical detection of a noisy signal. *Statistical Science* Vol. 23, No. 3, 354-364.
- DUNNINGTON, F. G. (1932). A Determination of e/m for an electron by a new deflection method. *Phys. Rev.* 43, 404-416.
- EFRON, B. (2005). Bayesians, frequentists, and scientists. *Journal of the American Statistical Association*, 100(469), 1-5.
- FRASER, D. A. S. (1991). Statistical inference: likelihood to significance. *J. Amer. Stat. Assoc.* **86**, 258-267.

- FRASER, D. A. S., REID, N. e WONG, A. C. M. (2004). Inference for bounded parameters. *Phys. Rev. D* **69** 033002.
- GALILEI, G. (1929-30). *Opere, da Lettera a Cristina di Lorena*, Vol. V, ed. Nazionale, Firenze.
- KLEIN, J. R. e ROODMAN, A. (2005). Blind analysis in nuclear and particle physics. *Annual Review of Nuclear and Particle Systems*, vol. 55, Issue 1, 141-163.
- LINNEMANN, J., LYONS, L. e REID, N. (2006). Report of the Banff Workshop, disponibile a: <http://www.birs.ca/workshops/2006/06w5054/report06w5054.pdf>.
- LISEO, B. (2008). *Introduzione alla Statistica Bayesiana*. Disponibile a: <http://3w.eco.uniroma1.it/utenti/liseo/didattica/bayes.pdf>.
- LYONS, L. (2008). Open statistical issues in particle physics. *Ann. Appl. Statist.* **2** 887915.
- MANDELKERN, M. (2002). Setting confidence intervals for bounded parameters (with discussion). *Statist. Sci.* **17** 149172.
- PACE, L. e SALVAN, A. (1996). *Teoria della Statistica: Metodi, Modelli, Approssimazioni Asintotiche*. Cedam, Padova.
- PACE, L. e SALVAN, A. (2001). *Introduzione alla Statistica II: Inferenza, Verosimiglianza e Modelli*. Cedam, Padova.
- PENROSE, R. (2005). *La Strada che Porta alla Realtà*, RCS Libri S.p.A., Milano.
- POPPER, K. R. (1957). *The Aim of Science*, trad. it. Lo scopo della scienza, in *Conoscenza oggettiva*, Armando Editore.

Ringraziamenti

Inserire i ringraziamenti al termine di questa tesi mi sembrava suonasse un po' eccessivo per un traguardo che è tutt'altro che una fine. Ho pensato però che se il traguardo è modesto, gli aiuti non lo sono stati, ed è più che giusto attribuire, almeno alle persone più importanti, i meriti di ciò che di buono ho raggiunto.

Sono profondamente convinta che ciascuno di noi sarebbe poca cosa senza le persone che ha incontrato lungo la strada. Io ho avuto la fortuna di conoscerne di straordinarie. Senza il loro supporto, la loro pazienza, il loro affetto, la loro generosità, la loro intelligenza e buonsenso, non sarei arrivata qui. O ci sarei arrivata infinitamente più povera.

Cercherò di ringraziare brevemente (ma non credo che ci riuscirò), in ordine di 'apparizione' perché sarebbe impossibile definirne uno di importanza, quelle che hanno lasciato un segno più profondo.

Ringrazio i miei genitori per aver sempre fatto tutto il possibile per darmi tutto ciò di cui potessi aver bisogno, materialmente e moralmente, per avermi dato la possibilità di studiare nel modo più sereno, per la pazienza che hanno avuto con una figlia quasi sempre assente negli ultimi anni, e ancor più nell'ultimo.

Ringrazio Vittoria, amica da tempi immemorabili, pur avendo intrapreso, da molti anni ormai, strade completamente diverse, non siamo mai state davvero lontane.

Ringrazio Maria, la mia nichilista, antiscientifica, americanofila, giramondo ex compagna di banco: anche se a volte mi chiedo cosa abbiamo in comune,

resta una delle persone che abbia più care al mondo. Da Sacramento a New York, da Barcellona a Stoccolma, non mi ha mai fatto mancare la sua presenza. Mi ha ricordato che ci si può anche vedere venti giorni in tre anni, ma se c'è un rapporto vero e un po' di impegno, questo non ha poi molta importanza, e mi ha ricordato tante volte quanto sia importante ciò che ho, ma anche che, comunque, lì fuori c'è un mondo. Potrei elencare decine di altri motivi, ma non mi sembra il luogo. Era però assolutamente necessario ringraziarla di cuore e... Hai visto che c'è il tuo nome in 'robe di matematica e di fisica'? Non capirò mai perché per tu me ne sia così grata, ma ho mantenuto la promessa!

Ringrazio Martina e Giuliana, che di quando in quando sono venute a trovarmi negli ultimi anni, mi ha sempre fatto un grande piacere.

Ringrazio la mia prof.ssa di lettere classiche del ginnasio, la prof.ssa Tentori, per averci insegnato con infinita pazienza e buona volontà il latino ed il greco, ma soprattutto l'italiano. E ringrazio il mio professore di filosofia, il prof. Cottini, perché rinunciando ad instillarci dell'arido e noioso nozionismo, fece una scelta molto più lungimirante: ci insegnò a pensare. Oltre ad essere stato infinitamente più divertente, ha lasciato a ciascuno di noi qualcosa che non avremmo potuto dimenticare e che avremmo utilizzato qualsiasi strada avessimo intrapreso nella vita.

Ringrazio le persone che ho incontrato trasferendomi a Padova, mi hanno fatto sentire sempre a casa e sono state parte di anni stupendi: Daniela per tutti i momenti passati insieme, per le serate a far notte sui progetti, per non avermi mai fatto mancare il suo aiuto ed i suoi consigli, per avermi sopportata nelle mie invettive contro la non-intelligenza artificiale (e per avermi poi dato una mano), per aver sopportato inoltre la mia testardaggine ed i miei "ci devo arrivare da sola" (con tutte le loro conseguenze); ma la ringrazio soprattutto per la lealtà, la generosità, per lo spirito di sacrificio, per esserci sempre stata, soprattutto quando era più difficile. Ringrazio la sua famiglia, che ormai sento un po' come la mia seconda famiglia, Flavia, Mario, Lisa, Matteo, che tutti i week-end mi hanno accolto davvero come

una nuova componente, con una generosità ed un calore che mai mi sarei aspettata: grazie, di cuore. Ringrazio la ‘compagnia’ che ha avuto un ruolo impagabile nel farmi ‘staccare la spina’ dopo tante giornate di studio: Diego, Roberto, Nadia, Manu e Marco.

Ringrazio Licia, per la comprensione da ex classicista dopo tante demoralizzanti lezioni di analisi o informatica, per le risate, la compagnia, perché è una delle persone con cui più è stato piacevole passare del tempo. Carlotta per l’entusiasmo, la vitalità, il buonumore sempre contagiosi. Illary per i tanti consigli, statistici e non, per la generosità e la disponibilità che ha sempre avuto nei miei confronti (e per aver cancellato un consistente, quanto inutile, numero di virgole nel Capitolo 1!).

Ringrazio la prof.ssa Salvan per la competenza, la pazienza e la gentilezza con la quale mi ha seguito nella stesura di questa tesi, e per avermi proposto l’argomento migliore possibile, ancora ne sono stupita. E ringrazio il prof. Sartori per la chiarezza e la gentilezza con le quali mi ha aiutato in quasi tutti, direi, i comandi R utilizzati per le stime ed i grafici.

Ringrazio infine tutti coloro che hanno sopportato il mio caratteraccio e che mi hanno dedicato un po’ del loro tempo, persone delle quali questa non potrà mai essere, di necessità, una lista esaustiva, ma per le quali ho sempre provato una profonda e sincera gratitudine, che spero di essere stata in grado di comunicare loro, malgrado tutto, nel corso degli anni.

