



UNIVERSITÀ DEGLI STUDI DI PADOVA

---

CORSO DI LAUREA MAGISTRALE  
IN SCIENZE STATISTICHE

TESI DI LAUREA

**METODI ASINTOTICI DI  
VEROSIMIGLIANZA PER  
ESPERIMENTI SEQUENZIALI**

RELATORE: PROF. NICOLA SARTORI

Dipartimento di Scienze Statistiche

CORRELATORE: PROF. ALESSANDRA SALVAN

Dipartimento di Scienze Statistiche

LAUREANDO: CAMILLA BISCO

---

ANNO ACCADEMICO 2011/2012



# Indice

<b>Introduzione</b>	<b>III</b>
<b>1 Esperimenti sequenziali a gruppi</b>	<b>1</b>
1.1 Introduzione . . . . .	1
1.2 Esperimenti comparativi . . . . .	3
1.3 Motivazioni e vantaggi . . . . .	6
1.4 Cenni storici . . . . .	8
<b>2 Aspetti teorici</b>	<b>13</b>
2.1 Introduzione . . . . .	13
2.2 Alcuni Richiami . . . . .	15
2.2.1 I test a numerosità campionaria fissa . . . . .	15
2.2.2 <i>Sequential Probability Ratio Test</i> (SPRT) . . . . .	18
2.2.3 Distribuzione congiunta di una successione di statistiche per risposte con distribuzione normale e varianza nota . . . . .	20
2.3 Modello Lineare normale . . . . .	23
2.4 Altri modelli parametrici . . . . .	26
2.5 Forma generale dei test sequenziali . . . . .	27
2.6 Incrementi disuguali nei livelli di informazione . . . . .	31
2.7 Test $t$ sequenziale . . . . .	32
2.8 Test unilaterali . . . . .	34
2.9 Alcuni test sequenziali . . . . .	36
2.9.1 Il test di Pocock . . . . .	36

2.9.2	Il test di O'Brien e Fleming . . . . .	38
2.9.3	I test di Wang e Tsiatis . . . . .	39
2.9.4	Proprietà e confronto . . . . .	41
<b>3</b>	<b>Il test triangolare di Whitehead</b>	<b>43</b>
3.1	Introduzione . . . . .	43
3.2	Caratteristiche principali dell'esperimento . . . . .	44
3.3	Test triangolare inverso e triangolare doppio . . . . .	50
3.4	Analisi dell'esperimento . . . . .	52
<b>4</b>	<b>Estensione del test di Whitehead</b>	<b>57</b>
4.1	Introduzione . . . . .	57
4.2	I metodi asintotici di ordine superiore . . . . .	58
4.2.1	Versioni modificate di $Z$ e $Z_p$ . . . . .	59
4.3	L'idea di base . . . . .	62
4.4	Struttura della ricerca . . . . .	62
4.5	Modelli e simulazioni . . . . .	65
4.5.1	Il modello normale con varianza ignota . . . . .	65
4.5.2	Il modello binomiale con parametro di interesse il <i>log odds ratio</i> . . . . .	71
4.5.3	Trasformazione arcoseno per il modello binomiale . . . . .	78
	<b>Considerazioni finali</b>	<b>85</b>
<b>A</b>	<b>Codice delle simulazioni</b>	<b>89</b>
A.1	Modello normale con varianza ignota . . . . .	89
A.2	Modello binomiale con parametro di interesse il <i>log odds ratio</i> . . . . .	96
A.3	Modello binomiale con trasformazione arcoseno . . . . .	100
	<b>Riferimenti bibliografici</b>	<b>103</b>

# Introduzione

La teoria classica del disegno di esperimenti è principalmente basata su metodi a numerosità campionaria fissata. In questo tipo di analisi la numerosità campionaria è predeterminata e i dati vengono analizzati una sola volta. In alternativa a tali metodi, è possibile utilizzare un'analisi di tipo sequenziale.

Quando in un esperimento si adotta una procedura sequenziale, i dati sono monitorati man mano che vengono raccolti. Lo scopo è quello di trarre informazione il prima possibile, per poter anticipare i tempi in caso di scelte o interventi di modifica al piano sperimentale. In generale è auspicabile un monitoraggio continuo dei dati, ma non sempre è realizzabile. Solitamente infatti i dati vengono analizzati a intervalli periodici nel corso dell'esperimento. Per questo si parla di metodi sequenziali a gruppi, poichè ad ogni analisi viene esaminato un gruppo di osservazioni.

Nell'ambito dell'analisi sequenziale, molte sono state le procedure proposte e le loro evoluzioni nel corso degli anni. In questa tesi si è focalizzata l'attenzione sul test triangolare di Whitehead (1997), basato su quantità di verosimiglianza. In particolare, il test di Whitehead utilizza come quantità chiave l'informazione parziale e la funzione *score*, entrambi calcolati rispetto al parametro di interesse,  $\theta$ , utilizzando una distribuzione di riferimento normale, almeno approssimativamente.

Il vantaggio derivante dall'utilizzo di metodi sequenziali di verifica d'ipotesi, rispetto ai metodi a numerosità campionaria fissata, è la riduzione della numerosità campionaria necessaria per garantire una data significatività e potenza. Talvolta i test sequenziali possono restituire evidenza empirica

anche avvalendosi di poche osservazioni. È immediato osservare che quando questo accade, i risultati asintotici possono essere carenti.

Il contributo che si desidera apportare con questa tesi è quello di modificare il test di Whitehead allo scopo di affinare l'approssimazione asintotica nei casi di bassa numerosità campionaria dei gruppi quando il parametro di interesse è scalare. In questo modo, si dovrebbe migliorare l'accuratezza nel controllo dell'errore di primo tipo effettivo. La modifica del test riguarda la sostituzione della statistica test utilizzata con una statistica test ad essa equivalente, ossia con la stessa distribuzione asintotica, almeno al primo ordine. Sono state proposte due alternative.

La prima alternativa sostituisce alla funzione *score* la versione unilaterale del test log rapporto di verosimiglianza moltiplicata per la radice dell'informazione parziale. A sostegno di ciò, lo stesso Whitehead (1997, Par. 3.9.1) spiega che è possibile condurre il test triangolare utilizzando il test di Wald o il test log rapporto di verosimiglianza al posto della *score*. Per grandi campioni infatti i tre test sono equivalenti, ma è possibile che in alcuni casi uno dei tre sia più accurato degli altri.

La seconda alternativa propone di applicare i metodi asintotici di ordine superiore. Al posto della *score* è stata utilizzata una versione modificata della radice con segno del log rapporto di verosimiglianza moltiplicata per la radice dell'informazione parziale. La convergenza alla distribuzione asintotica della versione modificata è infatti più veloce, e dovrebbe garantire risultati migliori in termini di accuratezza nei casi di bassa numerosità campionaria.

L'adeguatezza e l'accuratezza delle alternative proposte è stata successivamente valutata tramite simulazioni, in modelli sia per dati continui che discreti.

Il Capitolo 1 introduce gli esperimenti sequenziali illustrandone le caratteristiche principali e la loro evoluzione storica. Il Capitolo 2 è un capitolo di rassegna che descrive gli strumenti e le metodologie utilizzate negli esperimenti sequenziali. Vengono inoltre introdotti alcuni tra i principali test sequenziali. Il Capitolo 3 focalizza l'attenzione sul test triangolare di White-

---

head, spiegandone i fondamenti teorici e il funzionamento. Infine, il Capitolo 4, dopo alcuni richiami sui metodi asintotici di ordine superiore, ha lo scopo di presentare i risultati degli studi di simulazione svolti per valutare la bontà delle variazioni apportate al test di Whitehead. Nell'Appendice è riportato il codice utilizzato nelle simulazioni.





# Capitolo 1

## Esperimenti sequenziali a gruppi

### 1.1 Introduzione

Un esperimento è condotto sequenzialmente quando i dati vengono sottoposti ad analisi intermedie, man mano che sono disponibili. Il termine “a gruppi” indica che le analisi vengono svolte al raggiungimento di un insieme di  $m$  osservazioni aggiuntive. Il disegno sequenziale a gruppi è particolarmente indicato quando i dati vengono raccolti gradualmente nell’arco di un periodo di tempo; ad ogni analisi si analizzano i gruppi di risposte osservate fino a quel momento e si decide se continuare o terminare l’esperimento in base ad una regola di arresto definita in fase di pianificazione dello studio. L’obiettivo dell’esperimento è di verificare delle ipotesi di interesse. Non appena si ottiene evidenza empirica in favore di una delle due ipotesi tale da rendere superflua la continuazione dell’esperimento, è possibile terminare l’esperimento anticipatamente. La regola di arresto è quindi strettamente collegata al problema di verifica d’ipotesi, e, come si vedrà, l’esito dell’esperimento può essere influenzato dal tipo di regola scelta. Non è sempre necessario che le prove siano svolte ad uguali intervalli di tempo, né che siano pianificate preventivamente.

Contrapposta all’analisi sequenziale è l’analisi a numerosità campionaria fissata, che analizza i dati una sola volta, quando l’esperimento è concluso.

L'esperimento si considera concluso, ad esempio, al raggiungimento di una certa numerosità campionaria o di un certo intervallo temporale, a seconda degli obiettivi e della ricerca. Se in un esperimento sequenziale si ha bassa probabilità di conclusione anticipata, o la possibilità di una modifica del disegno sperimentale in corso d'opera è remota, allora l'analisi a campione fisso è particolarmente indicata. Altrimenti è preferibile un approccio a prove sequenziali.

I disegni sequenziali trovano applicazione in qualsiasi campo permetta la suddivisione dell'esperimento in fasi e in cui l'arco temporale di indagine abbia una durata tale da consentire analisi intervallari. Tali disegni si adattano molto bene agli esperimenti in ambito medico, poiché i dati vengono raccolti gradualmente e spesso per periodi di tempo prolungato. Inoltre, ragioni etiche rilevanti spingono verso l'obiettivo di trovare una conclusione valida il prima possibile. Questo sarà l'ambito di studio sul quale ci si concentrerà. I riferimenti principali sono i volumi di Jennison e Turnbull (2000) e Whitehead (1997). I metodi considerati saranno principalmente quelli finalizzati al confronto di due terapie, solitamente una innovativa, **sperimentale**, e una già in uso, **di controllo**. Con un esperimento comparativo si desidera rispondere a domande che riguardano la superiorità o inferiorità del nuovo trattamento, la presenza di effetti collaterali minori, o se un nuovo trattamento più economico non è statisticamente diverso da quello di controllo. È possibile avvalersi di analisi bilaterali, nelle quali si confronta l'ipotesi nulla di uguaglianza dei due trattamenti, contro l'ipotesi alternativa di differenza tra i trattamenti in entrambi i sensi: sperimentale migliore o peggiore del controllo. Alcune volte sono più adatte analisi unilaterali nelle quali l'ipotesi alternativa di disuguaglianza tra i due trattamenti considera solo superiorità o solo inferiorità di un trattamento rispetto all'altro. Ovviamente, maggiore è lo svantaggio che deriva dall'impiego di un trattamento inferiore, maggiore è il beneficio che è possibile ottenere dagli esperimenti a prove sequenziali, soprattutto nel caso di arresto anticipato.

Gli esperimenti sequenziali che verranno illustrati in seguito sono normal-

mente inseriti in un articolato percorso di ricerca che comprende varie prove ed analisi. Per entrare a far parte di un iter completo di esperimenti, ed eventualmente essere accettato dalla comunità medica, un nuovo trattamento o farmaco deve avere alle spalle ricerche che ne dimostrino le potenzialità. Il punto di partenza è un vantaggio inizialmente solo teorico, che costituisce la base degli studi su animali o prove di laboratorio. Se sembra valerne la pena, la nuova terapia viene seriamente presa in considerazione e valutata, in un processo che convenzionalmente viene suddiviso in fasi. Ad esempio, Whitehead (1997, Paragrafo 1.2), ne identifica quattro. La **fase 1** riguarda esperimenti esplorativi, come prove tossicologiche. Un obiettivo potrebbe essere quello di trovare la dose giusta di farmaco da somministrare per evitare effetti collaterali inaccettabili. Normalmente viene condotta su un numero limitato di pazienti sani. Durante la **fase 2**, vengono condotte indagini pilota di dimensione ridotta, per valutare sia l'efficacia che la sicurezza dei trattamenti da somministrare. La **fase 3** è quella determinante, in cui viene deciso se il farmaco o trattamento apporta effettivamente un miglioramento. Viene condotto un esperimento su pazienti che soffrono della malattia in questione, i quali vengono casualmente assegnati o al nuovo trattamento o a quello corrente (o placebo). In questa fase comparativa, l'analisi e ogni aspetto statistico vengono tenuti maggiormente sotto controllo, perché porteranno ad una scelta definitiva. La **fase 4** si riferisce a test aggiuntivi di controllo in caso di promozione del trattamento sperimentale, principalmente per monitorare effetti collaterali che potrebbero manifestarsi dopo lungo periodo.

## 1.2 Caratteristiche generali degli esperimenti comparativi

In base ad alcuni criteri, vengono selezionati i pazienti che si ritengono idonei a partecipare all'esperimento e questi devono dare il loro consenso. Successivamente, ad ognuno viene assegnato uno dei due trattamenti. Lo schema

di assegnazione dev'essere tale da garantire che i gruppi assegnati a diversi trattamenti abbiano la stessa numerosità e all'incirca le stesse caratteristiche. Si richiede cioè che i gruppi siano bilanciati, in base al rischio inerente la patologia trattata ed eventualmente ad importanti fattori prognostici, come ad esempio sesso, età, possibilità terapeutiche, complicanze, fattori ambientali. Il metodo più semplice per associare pazienti e trattamenti è quello casuale che non considera nessuna caratteristica dei pazienti ammessi, ma non sempre è appropriato. In alcuni casi, si ritiene utile tener conto di alcune caratteristiche proprie di ciascun paziente (variabili di stratificazione). Combinando le variabili di stratificazione tra loro si ottengono delle categorie e i pazienti vengono classificati in base all'appartenenza ad una di esse. I pazienti appartenenti alle varie categorie vengono assegnati ai trattamenti tramite un disegno a blocchi permutati casualmente o con *biased coin designs*. Se le variabili di stratificazione sono numerose, è conveniente utilizzare diverse classificazioni, ognuna riferita ad una o più variabili di stratificazione e si assegnano i pazienti ai trattamenti con un sistema simile al *biased coin design*, ma un po' più complesso perché tiene conto di più classificazioni. Questo modo di procedere viene chiamato *minimization* perché ogni associazione è scelta in modo da uguagliare il prima possibile il numero dei due tipi di trattamenti all'interno di ogni categoria. Per descrizioni più dettagliate dei disegni a blocchi permutati casualmente, dei *biased coin designs* e della *minimization*, si rimanda al Capitolo 2 di Whitehead (1997).

È importante che l'omogeneità tra i vari gruppi sia raggiunta il prima possibile e in seguito mantenuta, poiché l'arresto dell'esperimento può avvenire in ogni momento. Esperimenti svolti con gruppi tra loro sbilanciati non verranno considerati attendibili, poiché un'eventuale disuguaglianza rilevata tra i trattamenti potrebbe derivare solamente dall'eterogeneità nei due gruppi e non da effettive differenze nell'efficacia dei trattamenti. Lo schema di assegnazione è pertanto alla base della credibilità di tutto l'esperimento.

È buona norma che lo schema di assegnazione dei pazienti ai diversi trattamenti sia preparato in maniera del tutto indipendente rispetto ai medici che

conducono l'esperimento, in modo che questi non conoscano il trattamento con cui hanno a che fare volta per volta. Spesso è definito dal dipartimento di statistica del centro di ricerca. Questo garantisce che l'esperimento venga condotto nel modo più neutrale possibile, e sia del tutto indipendente dal criterio di assegnazione. La conoscenza del trattamento assegnato potrebbe condurre ad alcune imparzialità nelle valutazioni del medico o ad effetti di tipo psicologico nei pazienti. Si parla perciò di esperimento a doppio cieco (*double-blind*) se né il paziente né il medico conoscono quale sia il trattamento assegnato, mentre di esperimento a singolo cieco (*single-blind*) se solamente il paziente non conosce il trattamento assegnatogli. Un'altra distinzione riguarda l'identificazione o meno del gruppo di appartenenza: si ha trattamento cieco (*treatment-blind*) se i partecipanti conoscono il gruppo di appartenenza, ma non sanno quale dei due sia quello trattato con il farmaco sperimentale, mentre si ha gruppo cieco (*subgroup-blind*) se i partecipanti non riescono nemmeno ad identificare i due gruppi. È necessario che gli studiosi responsabili dell'esperimento conoscano l'assegnazione di ciascun trattamento per poter intervenire nel caso di inconvenienti come la perdita delle pillole o l'assunzione di altri farmaci che possano interagire con la terapia in corso. Ovviamente non sempre è possibile nascondere il tipo di trattamento, ad esempio nel confronto tra una mastectomia totale (rimozione chirurgica del seno) e la chemioterapia (trattamento farmacologico) per il cancro al seno. Non c'è nessun modo per ovviare a questo tipo di problema e inevitabilmente si avranno interferenze di tipo psicologico e fisiologico.

Nella conduzione di un esperimento medico può succedere che si manifestino circostanze eccezionali, come morte di un paziente, peggioramento delle sue condizioni mentre è in attesa di essere sottoposto ad un intervento chirurgico o somministrazione del farmaco o della dose sbagliata. In un primo momento, tali eventi potrebbero sembrare casuali, indipendenti dal tipo di trattamento; spesso invece sono proprio manifestazioni di qualche effetto negativo imprevisto. Per esempio, la somministrazione di una dose di farmaco sbagliata, potrebbe sembrare un errore straordinario, ma potrebbe essere

indice del fatto che istruzioni troppo complicate inducono una maggiore percentuale di errore. In situazioni di questo tipo, l'approccio più sicuro è quello di includere i casi eccezionali nell'esperimento e considerarli alla stregua di tutti gli altri. L'esclusione dall'esperimento viene limitata solo a quei rari eventi che, con molta cautela e buone motivazioni, vengono considerati indipendenti dal tipo di trattamento; diversamente si rischierebbe di produrre risultati distorti, o almeno di incorrere in quest'accusa.

Un aspetto caratterizzante degli esperimenti sequenziali è la definizione della regola di arresto. Tale scelta può influenzare l'interpretazione del risultato o la durata dell'esperimento. Una volta deciso il criterio di arresto è possibile avere dei dubbi sul comportamento da adottare nella gestione dell'esperimento, per esempio nel caso di arresto prematuro o perché il criterio di arresto non si verifica fino alla fine dell'esperimento o perché si soddisfa il criterio di arresto ma con valori ambigui e si pensa che tale evento possa essere trascurato. Per tali motivi spesso gli esperimenti hanno una commissione di controllo della validità dell'esperimento. Questa commissione ha il diritto di terminare l'esperimento qualora lo ritenga necessario per valide ragioni, per esempio di tipo etico o perché il disegno si rivela inadatto, o di continuare l'esperimento nonostante l'indicazione contraria o di cambiare la regola di arresto durante il corso dell'esperimento. Esiste un inevitabile conflitto tra l'accuratezza dell'esperimento e gli obiettivi etici. Questo viene minimizzato se la regola di arresto è disegnata in modo tale da bloccare l'esperimento all'apparire di evidenti situazioni di svantaggio per i pazienti. Il disegno sequenziale è quindi in minor contrasto con i requisiti etici che non un esperimento a campione fisso.

### 1.3 Motivazioni e vantaggi

È naturale pensare che in qualsiasi esperimento in cui i dati si accumulano stabilmente durante un periodo di tempo, i risultati vengano monitorati man mano che le informazioni sono disponibili, con l'obiettivo di terminare

lo studio non appena si giunge a qualche evidenza o per modificare eventualmente il piano dell'esperimento. La cosa migliore sarebbe riuscire ad effettuare un monitoraggio continuo, ma questo spesso è impraticabile. In ogni caso la maggior parte dei benefici che si otterrebbero con monitoraggio continuo si possono raggiungere esaminando i dati anche cinque o dieci volte durante il corso dell'esperimento. Per questo si parla di gruppi sequenziali, (*group sequential*): ad ogni passo viene analizzato un gruppo di unità campionarie. Le molteplici ragioni che possono portare ad analisi intervallari durante la raccolta di dati si possono riassumere in etiche, amministrative ed economiche.

Negli esperimenti su persone, l'etica spinge a tenere sotto controllo costante i risultati, per evitare che i pazienti siano sottoposti a trattamenti poco sicuri, inferiori o non efficaci. Ragioni etiche spingono inoltre a concludere l'esperimento il prima possibile, in modo da poter utilizzare la terapia migliore, se viene rilevata superiorità, o per poter concentrare le energie su altri trattamenti promettenti in attesa di essere valutati, nel caso in cui non vengano rilevate differenze.

Le ragioni amministrative riguardano il bisogno di assicurare che l'esperimento sia condotto come pianificato, che i soggetti soddisfino i criteri di selezione e che i trattamenti siano applicati come indicato. A volte un controllo iniziale dei risultati potrebbe rivelare la presenza di problemi, dando la possibilità di rimediare prima di sprecare troppe risorse o denaro. Un altro aspetto è la necessità di controllare la validità delle assunzioni fatte nel pianificare l'esperimento; eventuali assunzioni inappropriate possono quindi essere corrette in tempo.

La spinta iniziale per lo sviluppo dei disegni sequenziali è stata indotta da ragioni economiche nell'ambito del controllo della qualità. Se un esperimento ha esito positivo, questo economicamente si traduce nel fatto che un nuovo prodotto può essere immediatamente sfruttato, mentre se ha esito negativo un arresto anticipato significa che non vengono sprecate risorse inutilmente. I metodi sequenziali tipicamente inducono ad un risparmio in termini di

numerosità campionaria, tempo e costi se confrontati con una procedura standard a campione fisso. La numerosità campionaria diventa una variabile casuale con una distribuzione che dipende dalla regola di arresto scelta, dalla vera differenza tra i due trattamenti e dalla distribuzione delle osservazioni.

L'approccio sequenziale apporta molti vantaggi ma non sempre è implementabile. La procedura di raccolta e analisi dei dati per essere affidabile deve essere pianificata anticipatamente, e questo a volte non è possibile.

## 1.4 Cenni storici

Di seguito, basandosi su quanto riportato nel volume di Jennison e Turnbull (2000, Paragrafo 1.3), verranno riassunte le tappe che hanno portato alla nascita e allo sviluppo degli esperimenti sequenziali. Nel corso della storia del disegno degli esperimenti, l'approccio ai metodi sequenziali è avvenuto in modo naturale. Le prime applicazioni formali iniziarono negli anni '20 nell'area del controllo statistico della qualità per prodotti manifatturieri. Ad esempio, Dodge e Roming (1929) definirono un piano a due stadi per dati a risposta binaria (difettoso o non difettoso). Il piano prevedeva sei parametri: la numerosità campionaria per ogni stadio ( $n_1$  e  $n_2$ ), soglie di accettazione ( $c_1$  e  $c_2$ ) e soglie di rifiuto ( $d_1$  e  $d_2$ ) dove  $d_1 > c_1 + 1$  e  $d_2 = c_2 + 1$ . Si prendeva un campione iniziale di  $n_1$  merci, se al suo interno le merci difettose erano in numero inferiore o uguale a  $c_1$  il lotto veniva accettato, se erano in numero maggiore o uguale a  $d_1$  il lotto si rifiutava e negli altri casi si rimandava la decisione al secondo campione di numerosità  $n_2$ . Alla seconda analisi il lotto veniva quindi accettato se il numero totale di merci difettose risultava minore o uguale a  $c_2$  e rifiutato se risultava maggiore uguale a  $d_2$ . Questo procedimento è facilmente generalizzabile in un piano multi-stadio, in cui sono svolte  $K$  analisi (Bartky, 1943). Il piano multi-stadio successivamente sviluppato dalla Columbia University Research Group (Freeman et al., 1948) venne utilizzato come base per definire gli standard militari di accettazione dei lotti di armi, MIL-STD-105E (1989).



La teoria moderna dell'analisi sequenziale si sviluppa a partire dal lavoro di Abraham Wald (1947) negli Stati Uniti e di George Barnard (1946) in Gran Bretagna. In particolare, l'influenza maggiore deriva dal lavoro di Wald (1947) relativamente al *sequential probability ratio test* (SPRT). Si suppone che le osservazioni derivino da una distribuzione con funzione di densità nota con parametro  $\theta$ . Nel test sequenziale, nella sua forma base in cui ipotesi nulla e alternativa hanno una forma semplice ( $H_0 : \theta = \theta_0$  contro  $H_1 : \theta = \theta_1$ ), le osservazioni vengono estratte casualmente in modo sequenziale finché il valore del rapporto di verosimiglianza resta compreso in un certo intervallo  $(a, b)$ ; altrimenti si blocca l'esperimento che avrà come conclusione l'ipotesi indicata. Le costanti  $a$  e  $b$  vengono scelte in modo da mantenere le probabilità di errore di I tipo (falso negativo) e di II tipo (falso positivo) approssimativamente uguali ai valori prescelti di  $\alpha$  e  $\beta$  rispettivamente. Questa procedura tipicamente richiede una numerosità campionaria inferiore rispetto ad un test a campione fissato. Infatti, Wald e Wolfowitz (1948) provarono che il test rapporto di verosimiglianza sequenziale (SPRT) gode della proprietà di avere la minima numerosità campionaria attesa (ASN: *Average Sample Number*) sia sotto  $H_0$  che sotto  $H_1$  tra tutti i test con probabilità di errore non superiori ad  $\alpha$  e  $\beta$ . Tuttavia, il test SPRT comporta alcuni svantaggi: innanzitutto non impone nessun tipo di vincolo alla numerosità campionaria e questo potrebbe provocare una distribuzione della numerosità campionaria asimmetrica con varianza elevata. Inoltre ASN sarà più alta quando  $\theta$  è diverso da  $\theta_0$  o  $\theta_1$ . A tal proposito, una semplice modifica discussa da Wald è di troncare il test SPRT al raggiungimento di una certa numerosità campionaria. Inoltre, nella letteratura successiva, sono state considerate barriere non lineari e non parallele, utilizzando valori critici  $a$  e  $b$  non costanti e dipendenti dalla numerosità campionaria cumulata  $mk$ . Al crescere di  $mk$  le barriere hanno la tendenza ad avvicinarsi finché  $a_{mK} = b_{mK}$ , dove  $mK$  è un limite superiore alla numerosità campionaria.

Come si è visto, i test sequenziali sono stati inizialmente adottati per risposte binarie. Procedure per risposte con distribuzione normale a due e tre

stadi sono state sviluppate negli anni '60 da Armitage e Schneiderman (1958), Schneiderman (1961), Dunnett (1961) e Roseberry e Gehan (1964). In questi lavori, vengono utilizzate tecniche di integrazione numerica per calcolare la distribuzione della numerosità campionaria attesa. Tali metodi sono ancora oggi uno strumento chiave nella costruzione di molti test sequenziali a gruppi e su di essi si basano le procedure per il calcolo del livello di significatività e degli intervalli di confidenza in tali test.

I lavori di Pocock (1977), O' Brien e Fleming (1979) e Lan e De Mets (1983) sono stati particolarmente influenti ed insieme formano il punto di partenza per le recenti ricerche metodologiche e la base per gli esperimenti clinici.

Pocock (1977) definì chiare linee guida per raggiungere assegnati livelli di significatività e potenza negli esperimenti sequenziali a gruppi con risposte normali. Dimostrò la versatilità dell'approccio provando che il livello di significatività ottenuto per test sequenziali con risposte normali può essere utilizzato in modo affidabile anche per altri tipi di risposte e situazioni. In un articolo apparso poco dopo il lavoro di Pocock, O' Brien e Fleming (1979) proposero una diversa classe di test sequenziali a gruppi, basati su una revisione del test SPRT troncato. La novità in questi test sta nella tendenza a limitare le possibilità di arresto nelle prime analisi, e, se viene raggiunto l'ultimo stadio di analisi, effettuano un test simile a quello di un campione a numerosità fissa. Ulteriori contributi di Slud e Wei (1982) e Lan e De Mets (1983) dimostrarono che i test sequenziali possono essere utilizzati anche quando la numerosità di ogni gruppo è diversa e imprevedibile, situazione molto comune negli esperimenti in ambito medico.

Anche se l'obiettivo principale dei test sequenziali a gruppi è quello di scegliere tra ipotesi specifiche, è possibile fare ulteriore inferenza una volta giunti al termine di un esperimento, come calcolare stime puntuali,  $p$ -value, stime intervallari. Poiché il numero di osservazioni in un test sequenziale varia con il risultato raggiunto, non è possibile definire in modo univoco un ordinamento dei punti dello spazio campionario in base al rapporto di ve-

rosimiglianza. Pertanto l'ordinamento dello spazio campionario sul quale si basano gli intervalli di confidenza e i  $p$ -value, può essere definito con una certa libertà e in letteratura si sono sviluppate svariate proposte a riguardo. In tali approcci uno dei problemi ricorrenti è il ruolo predominante della regola di arresto. Come abbiamo anticipato, è buona norma che in un esperimento sequenziale ci sia una commissione di controllo. L'eventuale decisione di terminare in anticipo o continuare un esperimento, può essere causa di forti contrasti e critiche. L'analisi statistica di tipo sequenziale non è ancora ben equipaggiata per gestire una regola di arresto mutevole nel corso dell'esperimento. Le ricerche più recenti hanno cercato di sviluppare metodi più flessibili da questo punto di vista, come lo *stochastic curtailment*, gli intervalli di confidenza ripetuti (Jennison e Turnbull, 1984, 1989) e approcci Bayesiani per il monitoraggio flessibile (Spiegelhalter, Freedman e Parmar, 1994).

Per un approfondimenti si vedano i volumi di Jennison e Turnbull (2000), Wetherill e Glazebrook (1986) e Ghosh e Sen (1991), in particolare il primo capitolo (Ghosh, 1991).



# Capitolo 2

## Aspetti teorici

### 2.1 Introduzione

Quando in un esperimento sequenziale i dati accumulati sono analizzati a determinati intervalli temporali, anziché in modo continuo, si parla di test sequenziali a gruppi. Il monitoraggio continuo richiede l'impiego di un grande carico di risorse, perciò l'introduzione di metodi più parsimoniosi come l'analisi a gruppi ha favorito un maggiore impiego dei metodi sequenziali.

L'esempio principale che verrà considerato è quello del confronto tra due trattamenti in cui sono definiti il massimo numero di gruppi di osservazioni,  $K$ , e la numerosità di ogni gruppo,  $m$ , per ogni trattamento. I pazienti vengono assegnati ai trattamenti in base a uno schema casuale, soggetti al vincolo che  $m$  pazienti ricevano uno dei due trattamenti in ogni gruppo; i dati accumulati vengono poi analizzati una volta raggiunte  $m$  risposte per ciascun trattamento.

L'obiettivo principale è effettuare un test d'ipotesi, dove l'ipotesi nulla è sempre  $H_0 : \theta = 0$  e indica uguaglianza dei due trattamenti, mentre l'ipotesi alternativa può cambiare in base agli scopi dell'esperimento.

Viene definito un parametro di interesse scalare,  $\theta$ , che riflette la differenza tra i due gruppi, o più comunemente una misura del vantaggio del trattamento sperimentale rispetto a quello di controllo. Le scelte di  $\theta$  più co-

muni sono ad esempio la differenza delle medie dei due gruppi o la differenza delle medie standardizzata nel caso di dati provenienti da una distribuzione normale, oppure nel caso di dati binari, la differenza o il rapporto tra le due probabilità di successo o il logaritmo del rapporto delle quote (*log odds ratio*).

Per ogni analisi  $k = 1, \dots, K$ , verrà calcolata una statistica test, che varia in base all'approccio che si decide di seguire, e viene opportunamente standardizzata in modo che abbia, almeno approssimativamente, distribuzione nulla normale standard. In base alla statistica test scelta, verranno definite opportune sequenze di valori, detti barriere, per determinare la regione di rifiuto in modo da mantenere una probabilità di errore di I tipo pari ad  $\alpha$  (esattamente o approssimativamente a seconda del modello) per l'esperimento nel suo complesso.

In generale, se si verifica il criterio di arresto, ovvero la statistica test ha un valore compreso nella regione di rifiuto, l'esperimento termina rifiutando  $H_0$ ; altrimenti continua e si ripete la verifica all'analisi successiva nello stesso modo, finchè non si arriva all'analisi dell'ultimo gruppo. La numerosità  $m$  di ogni gruppo, è determinata invece dalla potenza che il test deve raggiungere.

In questo capitolo, il Paragrafo 2 è dedicato a richiami sui test a numerosità campionaria fissata, sul test SPRT di Wald, e su alcuni risultati alla base dei test sequenziali. Successivamente, nei Paragrafi 3 e 4, vengono considerate procedure sequenziali applicate a diversi tipi di modelli, per arrivare ad una formulazione generale degli esperimenti sequenziali nei Paragrafi 5 e 6. Nel Paragrafo 7 viene riassunta la procedura da utilizzare nel caso di test sequenziali per risposte normali con varianza ignota, nel Paragrafo 8 vengono introdotti i test sequenziali unilaterali con particolare riferimento a quello di Emerson e Fleming (1989). Infine, nel Paragrafo 9 viene illustrato il funzionamento dei test sequenziali di Pocock (1977), O'Brien e Fleming (1979) e Wang e Tsatis (1987).

## 2.2 Alcuni Richiami

### 2.2.1 I test a numerosità campionaria fissa

Siano  $X_i$  e  $Y_i$ ,  $i = 1, 2, \dots, n$ , le variabili casuali rappresentanti le risposte dei soggetti assegnati a due trattamenti, uno di controllo, il cui parametro verrà indicato con il pedice  $C$  e uno sperimentale, il cui parametro sarà contraddistinto dal pedice  $E$ . Si supponga che siano indipendenti con distribuzione normale con varianza nota:  $X_i \sim N(\mu_C, \sigma^2)$ ,  $Y_i \sim N(\mu_E, \sigma^2)$ . Si consideri la verifica d'ipotesi  $H_0 : \mu_E = \mu_C$  contro l'alternativa bilaterale  $H_1 : \mu_E \neq \mu_C$ , con probabilità di errore di I tipo  $\alpha$  e potenza  $1 - \beta$  per  $\mu_E - \mu_C = \pm\delta$ . Se ad ogni trattamento sono assegnati complessivamente  $n$  soggetti, per verificare l'ipotesi si può utilizzare la statistica test standardizzata  $Z_{ep}$  che avrà la forma

$$Z_{ep} = \frac{1}{\sqrt{(2n\sigma^2)}} \left( \sum_{i=1}^n Y_i - \sum_{i=1}^n X_i \right) \quad (2.1)$$

con distribuzione  $N((\mu_E - \mu_C)\sqrt{n/(2\sigma^2)}, 1)$ . Si avrà quindi che  $Z_{ep} \sim N(0, 1)$  sotto  $H_0$  e il test rifiuta  $H_0$  se  $|Z_{ep}| > \Phi^{-1}(1 - \alpha/2)$ .

I requisiti di potenza per il test con livello  $\alpha$ , considerando che sotto l'ipotesi alternativa  $H_1 : \mu_E - \mu_C = \pm\delta$  si ha  $Z_{ep} \sim N(\pm\delta\sqrt{n/(2\sigma^2)}, 1)$ , sono soddisfatti se

$$\begin{aligned} 1 - \beta &= P\{|Z_{ep}| > \Phi^{-1}(1 - \alpha/2) - \delta\sqrt{n/2\sigma^2}\} \\ &= P\{Z_{ep} > \Phi^{-1}(1 - \frac{\alpha}{2}) - \delta\sqrt{\frac{n}{2\sigma^2}}\} + P\{Z_{ep} < -\Phi^{-1}(1 - \frac{\alpha}{2}) - \delta\sqrt{\frac{n}{2\sigma^2}}\} \\ &= 1 - \Phi(\Phi^{-1}(1 - \alpha/2) - \delta\sqrt{n/2\sigma^2}) + \Phi(-\Phi^{-1}(1 - \alpha/2) - \delta\sqrt{n/2\sigma^2}) \\ &= \Phi(-\Phi^{-1}(1 - \alpha/2) + \delta\sqrt{n/2\sigma^2}) + \Phi(-\Phi^{-1}(1 - \alpha/2) - \delta\sqrt{n/2\sigma^2}). \end{aligned}$$

Dato che la probabilità di rifiutare  $H_0 : \mu_E - \mu_C = 0$  in favore di  $\theta < 0$  quando il vero valore è  $\theta = \delta$  è trascurabile, si ha che  $1 - \beta \doteq \Phi(-\Phi^{-1}(1 - \alpha/2) + \delta\sqrt{n/2\sigma^2})$ .

Esplicitando  $n$  in questa equazione si ottiene la numerosità campionaria

che soddisfa i requisiti di potenza

$$n_{f,2} = (\Phi^{-1}(1 - \beta) + \Phi^{-1}(1 - \alpha/2))^2 2\sigma^2/\delta^2 \quad (2.2)$$

che è il numero di pazienti da assegnare ad ognuno dei due trattamenti. Perciò complessivamente i soggetti osservati nell'esperimento saranno  $2n_{f,2}$ . Se l'ipotesi alternativa fosse stata unilaterale, con gli stessi passaggi si ottiene:

$$n_{f,1} = (\Phi^{-1}(1 - \beta) + \Phi^{-1}(1 - \alpha))^2 2\sigma^2/\delta^2. \quad (2.3)$$

Si noti che lo stesso esempio può essere riscritto in termini di quantità di verosimiglianza utilizzando la notazione seguente.

Si consideri un modello statistico con generica funzione di densità  $f(x; \psi)$  dove il parametro  $\psi \in \Psi \subseteq \mathbb{R}^p$ , è suddiviso in due componenti,  $\psi = (\theta, \lambda)$  una di interesse,  $\theta$ , e una disturbo,  $\lambda$ . La componente sulla quale si desidera fare inferenza è  $\theta$  che ha dimensione 1, mentre la componente di disturbo ha dimensione  $p - 1$ . La funzione di log verosimiglianza del modello è  $l(\psi)$ , e se si è interessati a fare inferenza sul solo parametro di interesse si utilizza la log verosimiglianza parziale, o log verosimiglianza profilo,  $l_P(\theta) = l(\hat{\psi}_\theta) = l(\theta, \hat{\lambda}_\theta)$ , dove  $\hat{\lambda}_\theta$  è la stima di massima verosimiglianza di  $\lambda$  considerando  $\theta$  fissato. Il vettore delle derivate parziali prime è  $l_*(\psi)$ . Di esso, l'elemento riferito solamente alla derivata rispetto a  $\theta$  si denota con  $l_\theta(\psi)$  e similmente per gli altri parametri. Si definisce, per semplificare la notazione,  $S(\theta) = l_\theta(\theta, \hat{\lambda}_\theta)$ , la *score* rispetto a  $\theta$  calcolata in  $(\theta, \hat{\lambda}_\theta)$ . La matrice di informazione osservata, ossia la matrice delle derivate seconde rispetto a  $\psi$  cambiate di segno, è  $j(\psi)$  e ha dimensione  $p \times p$ . Facendone il valore atteso si ottiene la matrice di informazione attesa o di Fisher:  $i(\psi)$ . Nel seguito verrà utilizzato il seguente risultato asintotico, sotto condizioni di regolarità,

$$\hat{\psi} \sim N_p(\psi, i(\psi)^{-1}),$$

dove il simbolo  $\sim$  indica che la distribuzione è approssimata, e  $i(\psi)$  può



essere sostituita dall'informazione osservata  $j(\psi)$  o  $j(\hat{\psi})$  (cfr. Pace e Salvan, 2001, Paragrafo 6.4). Nel caso si desideri, come negli esempi seguenti, fare inferenza sul solo parametro  $\theta$ , si utilizza l'informazione attesa parziale:

$$V(\theta, \lambda) = i^{\theta\theta}(\theta, \lambda)^{-1} = i_{\theta\theta}(\theta, \lambda) - \{i_{\theta\lambda}(\theta, \lambda)\} \{i_{\lambda\lambda}(\theta, \lambda)\}^{-1} \{i_{\theta\lambda}(\theta, \lambda)\}^T,$$

dove ad esempio,  $i_{\theta,\lambda}$  è il blocco relativo a  $\theta$  e  $\lambda$  della matrice  $i(\psi)$ , mentre  $i^{\theta\lambda}$  è l'analogo blocco di  $i(\psi)^{-1}$ . Si ha che  $V(\theta, \lambda)$  dipende sia dal parametro di interesse che da quello di disturbo. Di seguito verrà utilizzata nella forma  $V(\theta) = i^{\theta\theta}(\theta, \hat{\lambda}_\theta)^{-1}$ , dove al parametro di disturbo è stata sostituita la sua stima vincolata. La quantità analoga a  $V(\theta)$ , ma riferita all'informazione osservata  $j(\hat{\psi}_\theta)$  verrà indicata con  $I(\theta)$ .

Si ha quindi che  $\hat{\theta}$  avrà distribuzione asintotica  $N(\theta, V(\theta)^{-1})$ , e anche qui il risultato non cambia se a  $V(\theta)$  viene sostituito  $I(\theta)$  o  $I(\hat{\theta})$  (si veda Pace e Salvan 1996, Paragrafo 4.6). Vengono inoltre definiti il test log rapporto di verosimiglianza, il test *score* e il test di Wald che sono, in ordine,

$$\begin{aligned} W &= 2(l(\hat{\psi}) - l(\psi)), \\ W_u &= l_\psi(\psi)^T i(\psi)^{-1} l_\psi(\psi), \\ W_e &= (\hat{\psi} - \psi)^T i(\psi) (\hat{\psi} - \psi) \end{aligned}$$

e hanno distribuzione asintotica nulla  $\chi_p^2$ . Nel caso in cui il parametro  $\psi$  sia scalare, tali test vengono considerati nella loro versione direzionale, cioè

$$\begin{aligned} Z &= \text{sgn}(\hat{\psi} - \psi) \sqrt{2(l(\hat{\psi}) - l(\psi))}, \\ Z_u &= l_\psi(\psi) (i(\psi))^{-1/2}, \\ Z_e &= (\hat{\psi} - \psi) (i(\psi))^{1/2}, \end{aligned}$$

con distribuzione asintotica nulla  $N(0,1)$ . In presenza di parametri di disturbo, ma con  $\theta$  scalare, verranno considerate le analoghe versioni profilo,

ossia

$$\begin{aligned} Z_p &= \text{sgn}(\hat{\theta} - \theta) \sqrt{2(l_P(\hat{\theta}) - l_P(\theta))}, \\ Z_{up} &= S(\theta) V(\theta)^{-1/2}, \\ Z_{ep} &= (\hat{\theta} - \theta) V(\theta)^{1/2}, \end{aligned}$$

aventi distribuzione asintotica nulla  $N(0,1)$ . Per ulteriori approfondimenti si rimanda a Pace e Salvan (1996, Paragrafi 3.5 e 4.6).

Tornando all'esempio con risposte aventi distribuzione normale, si può ora notare che  $Z_{ep}$  nella (2.1) è il test di Wald nella versione profilo con  $V(\theta) = I(\theta) = i^{\theta\theta}(\theta)^{-1} = n/(2\sigma^2)$ . Sotto  $H_0$  ha distribuzione esatta  $N(0,1)$  e sotto  $H_1$  ha distribuzione esatta  $N(\theta\sqrt{V(\theta)}, 1)$ . Si noti che il requisito di potenza (2.2) può essere riscritto in termini di informazione attesa, essendo questa proporzionale alla numerosità campionaria, ottenendo

$$V(\theta) = (\Phi^{-1}(1 - \beta) + \Phi^{-1}(1 - \alpha/2))^2 / \delta^2,$$

che rappresenta il livello di informazione massima necessaria per raggiungere i requisiti di potenza.

### 2.2.2 *Sequential Probability Ratio Test (SPRT)*

Nell'ambito di problemi di verifica d'ipotesi semplici, il SPRT di Wald viene utilizzato qualora si voglia minimizzare il numero atteso di risposte da osservare. In questi casi infatti un test a campione fisso potrebbe non essere una buona scelta. Siano  $x_1, \dots, x_n$ , osservazioni provenienti dalle corrispondenti variabili aleatorie  $X_1, \dots, X_n$ , indipendenti ed identicamente distribuite (i.i.d.) con generica funzione di densità  $f(x; \psi)$  per  $X_i = x$ . Si supponga di voler verificare  $H_0 : \psi = \psi_0$  contro l'alternativa  $H_1 : \psi = \psi_1$ .

Il test SPRT definisce la regione di accettazione e rifiuto dell'ipotesi nulla

sulla base di due soglie  $A$  e  $B$  e del rapporto di verosimiglianza

$$\gamma_n = \prod_{i=1}^n f(X_i; \psi_1) / f(X_i; \psi_0).$$

dove  $f(X_i; \psi_1)$  e  $f(X_i; \psi_0)$  indicano la funzione di densità sotto l'ipotesi alternativa e sotto l'ipotesi nulla rispettivamente.

Se  $\gamma_n \leq A$  allora si accetta  $H_0$ , se  $\gamma_n \geq B$  si accetta  $H_1$  e se  $A < \gamma_n < B$  si passa ad analizzare una risposta successiva poiché non si verifica la regola di arresto.

Tale descrizione vale anche nel caso di osservazioni non i.i.d., basta considerare  $\gamma_n$  come il rapporto di verosimiglianza delle distribuzioni congiunte di  $(X_1, \dots, X_n)$  sotto  $H_0$  e sotto  $H_1$ .

Si definisca  $P_i(Q_i)$  la probabilità sotto  $\psi_i$  che si verifichi  $Q_i$ ; per qualsiasi insieme di osservazioni indipendenti  $(x_1, \dots, x_n)$  tale che  $\gamma_n$  sia nella zona di accettazione di  $H_1$ , si ha che  $\prod_{i=1}^n (f(X_i; \psi_1) / f(X_i; \psi_0)) \geq B$ , perciò

$$P_1(Q_1) \geq BP_0(Q_1). \quad (2.4)$$

In modo analogo, per qualsiasi insieme di osservazioni tale che  $\gamma_n$  sia nella zona di accettazione di  $H_0$ , si ha che  $\prod_{i=1}^n (f(X_i; \psi_1) / f(X_i; \psi_0)) \leq A$ , perciò

$$P_1(Q_0) \leq AP_0(Q_0). \quad (2.5)$$

Le probabilità di errore di I e II tipo possono essere definite come  $\alpha = P_0(Q_1)$  e  $\beta = P_1(Q_0)$  ed essendo  $Q_0$  e  $Q_1$  disgiunti, si ha:  $P_0(Q_0) = 1 - \alpha$  e  $P_1(Q_1) = 1 - \beta$ . Ora, dalle definizioni di errore di I e II tipo date e dalle equazioni (2.4) e (2.5) possiamo derivare due importanti disuguaglianze che devono essere soddisfatte dai valori  $A$  e  $B$

$$\begin{aligned} A &\geq \beta / (1 - \alpha) \\ B &\leq (1 - \beta) / \alpha. \end{aligned}$$

Tali disequaglianze diventano approssimativamente delle uguaglianze nel caso in cui le soglie  $A$  e  $B$  vengono superate di poco da  $\gamma_n$  quando termina il campionamento. Se nel test d'ipotesi si desidera fare in modo che le probabilità di errore di primo e secondo tipo non siano superiori ad  $\alpha$  e  $\beta$  rispettivamente, allora le soglie approssimate utilizzate sono  $A = \beta/(1 - \alpha)$  e  $B = (1 - \beta)/\alpha$ .

Il test SPRT è un test ottimo, ossia, se le osservazioni sono i.i.d. minimizza la numerosità campionaria attesa (*Average Sample Number*, ASN) per  $\psi = \psi_0$  e  $\psi = \psi_1$  tra tutti i test con probabilità di errore di primo e secondo tipo non superiori ad  $\alpha$  e  $\beta$ .

In varie situazioni pratiche si potrebbe essere interessati a verificare un'ipotesi composita del tipo  $H_0 : \psi \leq \psi_0$  contro  $H_1 : \psi \geq \psi_1$  ponendo il vincolo  $\alpha(\psi) \leq \alpha$  per tutti i  $\psi \leq \psi_0$  e  $\beta(\psi) \leq \beta$  per tutti i  $\psi \geq \psi_1$ . Se la distribuzione congiunta di  $(X_1, \dots, X_n)$  ha un rapporto di verosimiglianza monotono per ogni  $n$ , allora si può dimostrare che il test SPRT resta appropriato, ma non soddisfa nessuna proprietà di ottimalità nei casi in cui  $\psi \neq \psi_0, \psi_1$ . In ogni caso, spesso risulta che l'ASN del SPRT è minore della numerosità campionaria necessaria per un test corrispondente a campione fisso per ogni  $\psi$ . Per dimostrazioni e approfondimenti si vedano Wald (1945) e Kotz (2006, pp. 7607-7614).

### 2.2.3 Distribuzione congiunta di una successione di statistiche per risposte con distribuzione normale e varianza nota

Si supponga ora di utilizzare lo stesso modello del Paragrafo 2.2.1 per condurre il medesimo esperimento ma in modo sequenziale. L'obiettivo è sempre quello di verificare se l'uguaglianza dei due trattamenti è plausibile oppure no. La statistica utilizzata,  $Z_{ep,k}$ , avrà la forma (2.1), ma si avrà a che fare con una successione  $(Z_{ep,1}, \dots, Z_{ep,K})$  dove ad ogni analisi si utilizzano i valori disponibili all'analisi precedente e le nuove osservazioni. Al posto di  $n$  si utilizza  $mk$  perché il numero di osservazioni varia in base al gruppo che

si sta considerando. Come visto per i test a numerosità campionaria fissata, vale il risultato esatto  $Z_{ep,k} \sim N(\theta\sqrt{V_k(\theta)}, 1)$ , perciò, dato che ogni  $Z_{ep,k}$  è combinazione lineare di normali indipendenti, la distribuzione congiunta delle statistiche  $(Z_{ep,1}, \dots, Z_{ep,K})$  è una normale multivariata. Resta da calcolare la covarianza tra le statistiche dei vari gruppi. Per farlo si utilizza il fatto che la distribuzione dello stimatore di massima verosimiglianza per il parametro di interesse è  $\hat{\theta}^{(k)} \sim N(\theta, V_k(\theta)^{-1})$  (in questo caso è un risultato esatto), e che  $Z_{ep,k} = \sqrt{V_k(\theta)}\hat{\theta}^{(k)}$ , dove ora  $\hat{\theta}^{(k)} = (\sum_{i=1}^{mk} Y_i - \sum_{i=1}^{mk} X_i)/mk$  e  $V_k(\theta) = mk/2\sigma^2$ , rappresentano lo stimatore di massima verosimiglianza e l'informazione attesa parziale per  $\theta$  utilizzando le osservazioni fino alla  $k$ -esima analisi. Si ha che, sotto  $H_0$ ,

$$\begin{aligned} \text{Cov}(Z_{ep,k_1}, Z_{ep,k_2}) &= \text{Cov}(\sqrt{V_{k_1}(\theta)}\hat{\theta}^{(k_1)}, \sqrt{V_{k_2}(\theta)}\hat{\theta}^{(k_2)}) \\ &= \sqrt{V_{k_1}(\theta)}\sqrt{V_{k_2}(\theta)} \text{Cov}(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) \\ &= \sqrt{V_{k_1}(\theta)}\sqrt{V_{k_2}(\theta)} \text{Var}(\hat{\theta}^{(k_2)}) \\ &= \sqrt{V_{k_1}(\theta)/V_{k_2}(\theta)}, \quad \text{per } 1 \leq k_1 \leq k_2 \leq K. \end{aligned} \tag{2.6}$$

L'uguaglianza  $\text{Cov}(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) = \text{Var}(\hat{\theta}^{(k_2)})$  sfrutta il fatto che  $\hat{\theta}^{(k_1)}$  è calcolato utilizzando un sottoinsieme delle osservazioni con le quali si calcola  $\hat{\theta}^{(k_2)}$  ossia, esplicitando,

$$\begin{aligned} \hat{\theta}^{(k_1)} &= \frac{1}{mk_1} \left( \sum_{i=1}^{mk_1} Y_i - \sum_{i=1}^{mk_1} X_i \right) \\ \hat{\theta}^{(k_2)} &= \frac{1}{mk_2} \left( \sum_{i=1}^{mk_1} Y_i - \sum_{i=1}^{mk_1} X_i + \sum_{i=1+mk_1}^{mk_2} Y_i - \sum_{i=1+mk_1}^{mk_2} X_i \right) \\ &= \frac{1}{mk_2} \left( mk_1 \hat{\theta}^{(k_1)} + \sum_{i=1+mk_1}^{mk_2} (Y_i - X_i) \right) \end{aligned}$$

risulta

$$\begin{aligned} \text{Cov}(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) &= \text{Cov}\left(\hat{\theta}^{(k_1)}, \frac{k_1}{k_2}\hat{\theta}^{(k_1)} + \frac{1}{mk_2} \sum_{i=1+mk_1}^{mk_2} (Y_i - X_i)\right) \\ &= \frac{k_1}{k_2} \text{Var}(\hat{\theta}^{(k_1)}) + 0 = \frac{2\sigma^2}{mk_2} \\ &= \text{Var}(\hat{\theta}^{(k_2)}). \end{aligned}$$

In alternativa alla statistica di Wald si potrebbero utilizzare la *score*, o la statistica log rapporto di verosimiglianza. In riferimento all'esempio del Paragrafo 2.2.1, si ha che sotto  $H_0$ ,  $S_k(0) = mk(\bar{Y} - \bar{X})/2\sigma^2 = \hat{\theta}^{(k)}V_k(\theta)$  e  $Z_{p,k} = \hat{\theta}^{(k)}\sqrt{V_k(\theta)}$ . I risultati riportati di seguito sono quindi esatti.

Per un generico  $\theta$  si ha  $S_k(\theta) \sim N(\theta V_k(\theta), V_k(\theta))$  e  $Z_{p,k} \sim N(\theta\sqrt{V_k(\theta)}, 1)$ . La covarianza tra le varie componenti per  $S_k(\theta)$  è pari a:

$$\begin{aligned} \text{Cov}(S_{k_1}, S_{k_1+k_2}) &= \text{Cov}(S_{k_1}, S_{k_1} + (S_{k_1+k_2} - S_{k_1})) \\ &= \text{Var}(S_{k_1}) \end{aligned} \tag{2.7}$$

mentre per  $Z_{p,k}$  vale lo stesso risultato (2.6). Si consideri inoltre la successione delle stime di massima verosimiglianza del parametro di interesse  $(\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(K)})$ . Con dati provenienti da una distribuzione normale  $N(\theta, \sigma^2)$ , con varianza nota, si ha che tutte le successioni considerate hanno distribuzione congiunta normale multivariata con

$$\begin{aligned} \hat{\theta}^{(k)} &\sim N(\theta, V_k(\theta)^{-1}) \quad \text{e} \quad \text{Cov}(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) = V_{k_2}(\theta)^{-1}, \\ Z_{ep,k} &\sim N(\theta\sqrt{V_k(\theta)}, 1) \quad \text{e} \quad \text{Cov}(Z_{k_1}, Z_{k_2}) = \sqrt{(V_{k_1}(\theta)/V_{k_2}(\theta))}, \\ Z_{p,k} &\sim N(\theta\sqrt{V_k(\theta)}, 1) \quad \text{e} \quad \text{Cov}(Z_{k_1}, Z_{k_2}) = \sqrt{(V_{k_1}(\theta)/V_{k_2}(\theta))}, \\ S_k(\theta) &\sim N(\theta V_k(\theta), V_k(\theta)) \quad \text{e} \quad \text{Cov}(S_{k_1}, S_{k_2}) = V_{k_1}(\theta), \end{aligned}$$

per  $k = 1, \dots, K$  e  $1 \leq k_1 \leq k_2 \leq K$ . Esprimere un test sequenziale a gruppi con barriere riferite a  $\hat{\theta}^{(k)}$ , o a una delle altre statistiche non altera le sue proprietà e, nel caso di risposte diverse da quelle normali a varianza nota, di solito si sceglie in favore del metodo computazionalmente meno oneroso.

Si noti che considerare le varie analisi e gli intervalli tra di esse in base a  $V_k(\theta)$  o a  $mk$  non produce alcuna differenza poiché le due quantità sono proporzionali.

I risultati sopra illustrati sono basati sul modello normale con varianza nota. In modelli più complessi è comunque possibile utilizzare tali statistiche test, e procedere in analogia agli esempi illustrati, utilizzando la loro distribuzione asintotica. Si utilizzeranno quindi distribuzioni approssimate e non più esatte. Quando, per via della complessità della situazione siamo obbligati a lavorare con distribuzioni approssimate, è auspicabile utilizzare la statistica test che fornisca la migliore approssimazione alla distribuzione normale.

In letteratura sono presenti esempi di utilizzo dei test sopra citati in modo sequenziale. Ad esempio  $Z_{ep,k}$  è utilizzato nel test Pocock, il test  $S_k(0)$  è utilizzato da Whitehead nel test triangolare. Per  $Z_{p,k}$  si veda Lai (1988) e per un esempio pratico Shih et al. (2010).

Quanto riportato finora può essere applicato ad altri modelli statistici. Nel seguito verrà esposto il caso di un modello lineare normale ed un modello parametrico generale. Come riferimento è stato utilizzato quanto riportato in Jennison e Turnbull (2000, Capitolo 3) e si considerano esperimenti sequenziali in generale, senza essere vincolati ai soli esperimenti comparativi.

## 2.3 Modello Lineare normale

Si ipotizzi di lavorare con osservazioni  $X_i$ ,  $i = 1, 2, \dots$ , indipendenti e normalmente distribuite con valore atteso linearmente dipendente dal vettore di parametri  $\beta = (\beta_1, \dots, \beta_p)^T$ . I dati sono raccolti ed esaminati in modo sequenziale ogni volta che si osservano  $m$  osservazioni aggiuntive con un massimo di  $K$  analisi. Ad ogni analisi i dati disponibili sono  $mk$ . Sia  $X^{(k)} = (X_1, \dots, X_{mk})^T$  il vettore delle osservazioni disponibili alla  $k$ -esima

analisi. Si assume che

$$X^{(k)} \sim N(D^{(k)}\beta, I_{mk}\sigma^2),$$

dove  $D^{(k)}$  denota la matrice del disegno per le prime  $mk$  osservazioni e  $I_{mk}$  è la matrice di identità  $mk \times mk$ . Lo stimatore dei minimi quadrati di  $\beta$  è calcolabile ad ogni analisi e per i primi  $k$  gruppi di osservazioni è

$$\hat{\beta}^{(k)} = (D^{(k)T}D^{(k)})^{-1}D^{(k)T}X^{(k)}$$

ed è noto che è normalmente distribuito con valore atteso pari a  $\beta$  e matrice di covarianza  $\text{Var}(\hat{\beta}^{(k)}) = (D^{(k)T}D^{(k)})^{-1}\sigma^2$ . L'informazione di Fisher  $i_k(\beta)$  è l'inversa di tale matrice,  $i_k(\beta) = \text{Var}(\hat{\beta}^{(k)})^{-1}$ . Poiché ogni  $\hat{\beta}^{(k)}$  è combinazione lineare di normali indipendenti, marginalmente  $\hat{\beta}^{(k)}$  ha distribuzione normale e il vettore  $(\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(k)})$  è una normale multivariata della quale si conosce la distribuzione di  $\hat{\beta}^{(k)}$  e si deve calcolare la covarianza tra i  $\hat{\beta}^{(k)}$ . Si ha che  $\text{Var}(\hat{\beta}^{(k)}) = (D^{(k)T}D^{(k)})^{-1}\sigma^2$  con  $(D^{(k)T}D^{(k)})$  che non dipende da  $\beta$ . Per  $k_1 < k_2$  i dati dell'analisi  $k_1$  sono un sottoinsieme di quelli dell'analisi  $k_2$ , perciò si può scrivere

$$\hat{\beta}^{(k_1)} = A^T X^{(k_2)}.$$

Dato che  $\hat{\beta}^{(k_1)}$  è uno stimatore non distorto per  $\beta$ , si ha che

$$E(A^T X^{(k_2)}) = A^T D^{(k_2)}\beta = \beta \quad \text{per ogni } \beta$$

ossia  $A^T D^{(k_2)} = I_p$ . Si ottiene quindi

$$\begin{aligned} \text{Cov}(\hat{\beta}^{(k_1)}, \hat{\beta}^{(k_2)}) &= \text{Cov}(A^T X^{(k_2)}, (D^{(k_2)T}D^{(k_2)})^{-1}D^{(k_2)T}X^{(k_2)}) \\ &= A^T \text{Var}(X^{(k_2)})D^{(k_2)}(D^{(k_2)T}D^{(k_2)})^{-1} \\ &= (D^{(k_2)T}D^{(k_2)})^{-1}\sigma^2 \\ &= \text{Var}(\hat{\beta}^{(k_2)}). \end{aligned}$$

Riassumendo, la distribuzione della successione delle stime dei parametri è



una distribuzione normale multivariata con

$$\begin{aligned}\hat{\beta}^{(k)} &\sim N_p(\beta, i_k(\beta)^{-1}) \quad 1 \leq k \leq K, \\ \text{Cov}(\hat{\beta}^{(k_1)}, \hat{\beta}^{(k_2)}) &= \text{Var}(\hat{\beta}^{(k_2)}), \quad 1 \leq k_1 \leq k_2 \leq K.\end{aligned}\tag{2.8}$$

Si supponga che il parametro di interesse in un esperimento sia ad esempio  $\theta = \beta_1$ , cioè il primo elemento del vettore  $\beta$ , e che si voglia verificare  $H_0 : \theta = 0$  contro un'alternativa bilaterale. In tale esperimento,  $\theta$  rappresenta l'effetto del trattamento, mentre gli altri elementi di  $\beta$  sono trattati come parametri di disturbo relativi a covariate che influenzano la risposta ma non sono di primario interesse. Il test sequenziale a gruppi per  $H_0 : \theta = 0$  si basa sulla successione di stime  $(\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(K)})$ . La varianza di  $\hat{\theta}^{(k)}$  alla  $k$ -esima analisi è l'elemento di posto  $(1, 1)$  della matrice di covarianza  $\text{Var}(\hat{\beta})^{(k)} = (D^{(k)T} D^{(k)})^{-1} \sigma^2$ , ed è il reciproco dell'informazione parziale di Fisher per  $\theta$ :

$$V_k(\theta) = \text{Var}(\hat{\theta}^{(k)})^{-1} = [\{(D^{(k)T} D^{(k)})^{-1}\}_{11} \sigma^2]^{-1}, \quad k = 1, \dots, K.$$

La statistica  $Z_{ep,k}$  per verificare  $H_0$  è quindi

$$Z_{ep,k} = \frac{\hat{\theta}^{(k)}}{\sqrt{\text{Var}(\hat{\theta}^{(k)})}} = \hat{\theta}^{(k)} \sqrt{V_k(\theta)} \quad k = 1, \dots, K.$$

Dato che il vettore  $(\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(K)})$  è una normale multivariata, ne segue che  $(\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(K)})$  è anch'esso normale multivariata con

$$\begin{aligned}E(\hat{\theta}^{(k)}) &= \theta, \quad 1 \leq k \leq K, \\ \text{Cov}(\hat{\theta}^{(k_1)}, \hat{\theta}^{(k_2)}) &= \text{Var}(\hat{\theta}^{(k_2)}), \quad 1 \leq k_1 \leq k_2 \leq K.\end{aligned}$$

Perciò si ha

$$\begin{aligned}(Z_{ep,1}, \dots, Z_{ep,K}) &\text{ normale multivariata,} \\ E(Z_{ep,k}) &= \theta \sqrt{V_k(\theta)} \quad 1 \leq k \leq K, \\ \text{Cov}(Z_{ep,k_1}, Z_{ep,k_2}) &= \sqrt{V_{k_1}(\theta)/V_{k_2}(\theta)}, \quad 1 \leq k_1 \leq k_2 \leq K.\end{aligned}\tag{2.9}$$

Analoghi risultati si ottengono per ipotesi del tipo  $H_0 : c^T \beta = \gamma$ . Il parametro di interesse è  $\theta = c^T \beta - \gamma$  con  $c$  vettore  $p \times 1$  e  $\gamma$  scalare. Gli stessi risultati del caso di incorrelazione valgono quando

$$X^{(k)} \sim N(D^{(k)}\beta, \Sigma^{(k)}\sigma^2),$$

con  $\Sigma^{(k)}$  matrice non diagonale e  $\sigma^2$  entrambi noti. La stima di massima verosimiglianza per  $\beta$  all'analisi  $k$  è

$$\hat{\beta}^{(k)} = (D^{(k)T} \Sigma^{(k)-1} D^{(k)})^{-1} D^{(k)T} \Sigma^{(k)-1} X^{(k)},$$

con distribuzione normale multivariata con valore atteso pari a  $\beta$  e varianza  $(D^{(k)T} \Sigma^{(k)-1} D^{(k)})^{-1} \sigma^2$ . La distribuzione congiunta di  $(\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(k)})$  è normale multivariata con valore atteso e matrice di covarianza (2.8).

## 2.4 Altri modelli parametrici

Si supponga di condurre un esperimento sequenziale a gruppi con osservazioni generate da una distribuzione qualsiasi con vettore di parametri  $\psi$  per i quali sia possibile calcolare le stime di massima verosimiglianza alle analisi  $k = 1, \dots, K$ . Grazie ad un risultato illustrato in Jennison e Turnbull (1997), sotto certe condizioni di regolarità del modello statistico e delle covariate dei valori osservati, la successione di stime  $(\hat{\psi}^{(1)}, \dots, \hat{\psi}^{(k)})$  è approssimativamente normale multivariata con

$$E(\hat{\psi}^{(k)}) = \psi, \quad 1 \leq k \leq K,$$

e

$$\text{Cov}(\hat{\psi}^{(k_1)}, \hat{\psi}^{(k_2)}) = \text{Var}(\hat{\psi}^{(k_2)}), \quad 1 \leq k_1 \leq k_2 \leq K,$$

quando la numerosità campionaria è sufficientemente elevata. La distribuzione congiunta per stime successive del vettore di parametri  $\psi$  è pertanto riconducibile a quella per osservazioni normali.

In Jennison e Turnbull (2000, Paragrafo 11.6), viene dimostrato che l'inversa della matrice di informazione di Fisher per  $\psi$  all'analisi  $k$  restituisce una stima consistente di  $\text{Var}(\hat{\psi}^{(k)})$ .

Ad esempio, se si desidera verificare un'ipotesi nulla del tipo  $H_0 : \theta = 0$  contro l'alternativa bilaterale utilizzando la statistica test  $Z_{ep,k}$ . Il parametro di interesse è  $\theta = c^T \psi - \gamma$ , con  $\gamma$  scalare e  $c$  vettore  $p \times 1$ . In base a quanto appena descritto, la successione  $(Z_{ep,1}, \dots, Z_{ep,K})$  ha distribuzione approssimata (2.9).

Le condizioni sotto le quali vale il precedente risultato dipendono dal particolare modello statistico che si sta utilizzando. In sintesi sono quelle richieste per applicare la teoria asintotica della verosimiglianza ad ogni singola analisi. La numerosità di ogni singolo gruppo di osservazioni,  $m$  è una variabile chiave, da determinare in base alle caratteristiche specifiche di ogni esperimento. È consigliabile sceglierla tenendo conto che la numerosità dei gruppi dovrebbe essere sufficientemente grande da garantire l'applicazione dei risultati asintotici a partire dalla prima analisi.

## 2.5 Forma generale dei test sequenziali

Utilizzando come riferimento Jennison e Turnbull (2000, Paragrafo 3.2) si fornisce ora una formulazione generale per i test sequenziali, che, con opportuni accorgimenti è possibile applicare a varie tipologie di dati, di test e problemi di verifica d'ipotesi. Il motivo sta nel fatto che le successioni di statistiche test fanno capo ad un'unica distribuzione congiunta comune che è la (2.9). Di seguito si considera il caso in cui il vettore di parametri generale  $\psi$  è suddiviso in una componente di interesse  $\theta$  di dimensione 1 e una componente di disturbo  $\lambda$  di dimensione  $p - 1$ . Si è fatta questa scelta poiché gli esperimenti comparativi che verranno considerati in seguito utilizzeranno la stessa suddivisione dei parametri. La statistica test scelta per la verifica d'ipotesi è  $Z_{ep,k}$ .

Si supponga di disporre di osservazioni suddivise in gruppi con identica numerosità provenienti da due tipi di trattamenti, uno di controllo, C e uno sperimentale, E. Il parametro sul quale si desidera fare inferenza è  $\theta$  che misura la differenza tra i due trattamenti. È richiesto di verificare l'ipotesi  $H_0 : \theta = 0$  contro l'ipotesi bilaterale  $H_1 : \theta = \pm\delta$ , con probabilità di errore di I tipo  $\alpha$  e potenza  $1 - \beta$  per  $\theta = \pm\delta$ . Sotto  $H_0$ , la distribuzione di  $(Z_{ep,1}, \dots, Z_{ep,K})$ , con  $1 \leq k_1 \leq k_2 \leq K$ , è approssimativamente:

$$\begin{aligned} & (Z_{ep,1}, \dots, Z_{ep,K}) \text{ normale multivariata,} \\ & E(Z_{ep,k}) = 0, \quad \text{per } k = 1, \dots, K, \quad \text{e} \\ & \text{Cov}(Z_{ep,k_1}, Z_{ep,k_2}) = \sqrt{V_{k_1}(\theta)/V_{k_2}(\theta)}, \end{aligned} \quad (2.10)$$

mentre, quando  $\theta = \pm\delta$ , la distribuzione approssimata diventa

$$\begin{aligned} & (Z_{ep,1}, \dots, Z_{ep,K}) \text{ normale multivariata,} \\ & E(Z_{ep,k}) = \pm\delta\sqrt{V_k(\theta)}, \quad \text{per } k = 1, \dots, K, \\ & \text{Cov}(Z_{ep,k_1}, Z_{ep,k_2}) = \sqrt{V_{k_1}(\theta)/V_{k_2}(\theta)}. \end{aligned} \quad (2.11)$$

Tali distribuzioni sono esatte nel caso di osservazioni normali con varianza nota. La probabilità di commettere un errore di I tipo è

$$\Pr_{\theta=0} \{ |Z_{ep,k}| \geq c_k \text{ per qualche } k = 1, \dots, K \}$$

valutata quando  $Z_{ep,k}$ ,  $k = 1, \dots, K$ , segue la distribuzione nulla (2.10). Le soglie  $c_k$  sono calcolate numericamente per fare in modo che questa probabilità sia uguale al valore di  $\alpha$  scelto. Per costruzione le barriere superiori ed inferiori del test saranno simmetriche. La potenza del test per  $\theta = \pm\delta$  è

$$\Pr_{\theta} \left\{ \bigcup_{k=1}^K (|Z_{ep,j}| < c_j \text{ per } j = 1, \dots, k-1 \text{ e } |Z_{ep,k}| \geq c_k) \right\} \quad (2.12)$$

valutata quando  $Z_{ep,k}$ ,  $k = 1, \dots, K$  segue la distribuzione (2.11). Dati  $K$ ,  $\alpha$ ,  $\beta$  e i valori critici  $(c_1, \dots, c_k)$  è possibile trovare un valore dell'informazione

che renda la probabilità (2.12) pari a  $1 - \beta$ . Tale valore è l'informazione massima richiesta per soddisfare i requisiti di potenza,  $V_{max}(\theta)$ . In base alla formulazione dell'informazione attesa parziale massima,  $V_{max}(\theta)$ , imponendo  $V_{max}(\theta) = V_K(\theta)$ , è possibile derivare i requisiti su  $mK$  come visto alla fine del Paragrafo 2.2.1. Si ricorda che  $V_K$  è il livello di informazione raggiunto nell'ultima analisi.

L'informazione massima necessaria, derivante dalla (2.12) ha la stessa formulazione dell'informazione di un test a numerosità campionaria fissata che verifica la stessa ipotesi,  $V(\theta)$ , dato che entrambe sono calcolate sullo stesso modello. Le due informazioni pertanto sono tra loro proporzionali:  $V_{max}(\theta) = RV(\theta)$ , dove  $R$  è una costante di proporzionalità maggiore di uno che dipende da  $K$ ,  $\alpha$ ,  $\beta$  e dal tipo di barriere che si sono scelte. Con livelli di informazione equispaziati e dati normali con varianza nota si ha

$$\begin{aligned} V_k(\theta) &= (k/K)V_{max}(\theta) = (k/K)RV(\theta) \\ &= \frac{kR}{K} \frac{\{\Phi^{-1}(1 - \alpha/2) + \Phi^{-1}(1 - \beta)\}^2}{\delta^2}, \quad k = 1, \dots, K, \end{aligned}$$

ed essendo  $V_k(\theta)$  proporzionale ad  $mk$ , è possibile ottenere il valore di  $m$  che soddisfa i requisiti di potenza. In fase di progettazione dell'esperimento è quindi possibile decidere in anticipo il livello di informazione  $V_k(\theta)$  al raggiungere del quale verrà effettuata la  $k$ -esima analisi.

Per alcuni test, come quelli di Pocock, O' Brien e Fleming, e Wang e Tsia-tis, che verranno illustrati in seguito, i valori di  $R$  sono tabulati e derivano da calcoli numerici. Pertanto, l'informazione massima necessaria affinché un test sequenziale soddisfi i requisiti di potenza è facilmente calcolabile utilizzando i valori di  $R$ ,  $K$ ,  $\alpha$ ,  $\beta$  e  $V(\theta)$ . Per altri test, come quello triangolare di Whitehead, l'informazione massima viene definita dal punto di intersezione delle barriere; nel Capitolo 3 verrà illustrato come calcolarla.

Una volta trovate le soglie  $c_k$ , la regola di arresto viene applicata come segue: per  $k = 1, \dots, K - 1$ , si calcola la statistica scelta, in questo caso la statistica standardizzata  $Z_{ep,k}$  e se  $|Z_{ep,k}| > c_k$  l'esperimento termina ri-

fiutando  $H_0$ , altrimenti continua analizzando il gruppo successivo. Quando  $Z_{ep,k} > c_k$  si ha evidenza di superiorità del trattamento sperimentale, e quando  $Z_{ep,k} < -c_k$  si ha evidenza di inferiorità del trattamento sperimentale. Se si arriva all'ultima analisi, la  $K$ -esima, l'esperimento termina e si accetta  $H_0$  se  $|Z_{ep,K}| < c_K$  altrimenti si rifiuta.

Si noti che  $V_k(\theta)$  viene utilizzato solamente in fase di programmazione dell'esperimento. Quando l'esperimento è in corso,  $V_k(\theta)$  viene sostituito dall'informazione osservata,  $I_k(\theta)$ . Nel corso dell'esperimento possono dunque esserci delle discrepanze tra i valori attesi e quelli osservati.

Un metodo generale per impostare un esperimento sequenziale a gruppi può essere riassunto in 5 fasi (cfr. Jennison e Turnbull, 2000, Paragrafo 3.2)

1. Identificare il modello statistico che descriva la distribuzione dei dati e il parametro di interesse  $\theta$  che deve avere valore pari a 0 sotto l'ipotesi nulla. Decidere la probabilità di errore di I tipo,  $\alpha$ . Scegliere  $\beta$  e  $\delta$  in modo tale che il test  $H_0 : \theta = 0$  abbia potenza  $1 - \beta$  per  $\theta = \pm\delta$ .
2. Scegliere il tipo di test, ovvero il tipo di barriere da utilizzare e il numero massimo di analisi  $K$ . Una scelta può essere fatta in base a quanto riportato nel Paragrafo 2.9.4.
3. Definire la statistica test che si utilizza e trovare l'espressione per l'informazione  $V_k(\theta)$  proporzionale al numero di osservazioni.
4. Trovare il fattore  $R$  adatto al tipo di test e al numero di analisi,  $K$ , quindi calcolare i livelli di informazione necessari per ogni analisi, ovvero  $V_k(\theta) = (k/K)RV(\theta)$ ,  $k = 1, \dots, K$ , dove  $V(\theta)$  è l'informazione di Fisher per un esperimento a campione fisso ad alternativa bilaterale. Disegnare l'esperimento ed il campionamento in modo che la numerosità campionaria permetta di raggiungere tali livelli di informazione.
5. Condurre l'esperimento applicando la regola di arresto spiegata precedentemente.

## 2.6 Incrementi disuguali nei livelli di informazione

L'assunzione di lavorare con gruppi di osservazioni con uguale numerosità, finora adottata trova scarsa applicazione a livello pratico. Può succedere che i livelli di informazione osservati durante il corso di un esperimento non siano equivalenti a quelli attesi, calcolati in fase di programmazione.

Le differenze tra i livelli di informazione osservati e quelli obiettivo conducono a deviazioni dalle probabilità di errore di I e II tipo pianificate. In particolare, deviazioni dal valore dell'informazione finale modificano la potenza effettiva, mentre deviazioni nella spaziatura dei livelli di informazioni hanno effetto sull'errore di I tipo. In situazioni di questo genere l'esperimento può essere condotto come programmato, e ad ogni analisi la verifica d'ipotesi verrà svolta utilizzando il valore dell'informazione effettivamente osservato anche se è diverso da quello obiettivo.

Questo modo di procedere, seppur approssimato, è considerato soddisfacente. In base a studi su dati simulati, le probabilità di errore di I e II tipo, calcolate quando l'informazione osservata è diversa da quella attesa, sono risultate molto vicine ai valori predeterminati. In particolare, Proschan, Follmann e Waclawiw (1992) hanno condotto uno studio per identificare l'incremento massimo nell'errore di I tipo sotto alcune restrizioni sulla numerosità osservata di ciascun gruppo. Si dimostra che il tasso di errore è robusto a deviazioni dai valori obiettivo, e raggiunge valori inaccettabili solo in casi estremi. La richiesta di avere almeno approssimativamente numerosità campionaria uguale tra i gruppi è molto limitante a livello pratico. Se si prevede di riscontrare ingenti ed inevitabili differenze nella numerosità dei gruppi osservati, è conveniente utilizzare approcci alternativi come l'*Error Spending Approach*, adatti a gestire diverse e imprevedibili numerosità campionarie. Per approfondire tale argomento si veda Whitehead (1997, Paragrafo 6.2).

## 2.7 Test $t$ sequenziale

Si consideri il modello lineare normale del Paragrafo 2.3, con la differenza che ora  $\sigma^2$  è ignoto. In questo caso la statistica di interesse sarà  $T_k$  che prende il nome di test  $t$ , e avrà distribuzione  $t$ -student con  $n - p$  gradi di libertà, dove  $p$  è il numero di elementi del vettore di parametri  $\beta$  sui quali si fa inferenza. Sia  $s_k^2$ ,  $k = 1, \dots, K$ , una stima non distorta di  $\sigma^2$  e  $D^{(1)}, \dots, D^{(K)}$ , una successione tale che  $\text{Var}(\hat{\beta}^{(k)}) = (D^{(k)T} D^{(k)})^{-1} \sigma^2$ . Si ipotizzi di voler fare inferenza su una combinazione lineare delle componenti del vettore di parametri  $\beta$ , cioè il parametro di interesse ha generica forma  $\theta = c^T \beta - \gamma$  di dimensione 1. Ad ogni analisi  $k$  la stima di  $\theta$  è

$$\hat{\theta}^{(k)} \sim N(\theta, \sigma^2 c^T (D^{(k)T} D^{(k)})^{-1} c).$$

Per verificare  $H_0 : \theta = 0$  contro un'alternativa bilaterale quando  $\sigma^2$  è noto si può utilizzare la statistica test  $Z_{ep,k}$  che ha la forma

$$Z_{ep,k} = \frac{\hat{\theta}^{(k)}}{\sqrt{\sigma^2 c^T (D^{(k)T} D^{(k)})^{-1} c}} = \hat{\theta}^{(k)} \sqrt{V_k(\theta)}.$$

Nel caso di varianza ignota si usa invece

$$T_k = \frac{\hat{\theta}^{(k)}}{\sqrt{s_k^2 c^T (D^{(k)T} D^{(k)})^{-1} c}}. \quad (2.13)$$

Sotto  $H_0$ ,  $T_k$  ha distribuzione  $t$ -student con  $mk - p$  gradi di libertà (in questo caso  $p=1$ ). È noto che per valori grandi di  $m$  la distribuzione  $t$ -student converge ad una distribuzione normale standard (si veda Pace e Salvani 2001, esempio 0.12). Sotto l'alternativa,  $H_1 : \theta = \pm\delta$ , avrà distribuzione  $T(mK - p; \delta / \sqrt{c^T (D^{(k)T} D^{(k)})^{-1} c s^2})$  dove  $T(v; \zeta)$  denota una distribuzione  $t$ -student con  $v$  gradi di libertà e parametro di non centralità  $\zeta$ . La probabilità di errore di I tipo per il test sequenziale nel suo complesso è data da:

$$\alpha = P(|T_1| > c_1) + \sum_{k=2}^K P(|T_k| > c_k, |T_1| \leq c_1, \dots, |T_{k-1}| \leq c_{k-1}) \quad (2.14)$$



Le soglie  $(c_1, \dots, c_K)$  devono soddisfare la (2.14) e vengono calcolate approssimando la distribuzione  $t$ -student con la normale. Se si effettua il test sequenziale utilizzando come valori critici le soglie  $(c_1, \dots, c_K)$  calcolate in questo modo, è possibile che il livello  $\alpha$  effettivo e la potenza raggiunta dal test siano leggermente diversi dal suo livello nominale. Tali valori critici saranno accurati se l'approssimazione alla normale è buona, ovvero quando si lavora con valori grandi di  $m$ .

Garantire la potenza pari a  $1 - \beta$  per uno specifico valore del parametro  $\theta$  è più difficile, dato che la potenza dipende da un valore stimato di  $\sigma^2$ . In analogia con il caso di varianza nota, stando a quanto riportato nel Paragrafo 2.5, ci si attende che la potenza di un test  $t$  sequenziale per una data alternativa, sia uguale a quello di un test  $t$  a campione fisso con informazione attesa parziale  $V(\theta)R = V_K(\theta)$ , cioè con varianza  $\text{Var}(\hat{\theta}) = R\text{Var}(\hat{\theta}^{(k)})$ . Perciò è sufficiente calcolare la potenza per un test a campione fisso con  $\text{Var}(\hat{\theta}) = R s^2 c^T (D^{(k)T} D^{(k)})^{-1} c$ .

La probabilità di rifiutare l'ipotesi nulla per valori molto grandi in valore assoluto e negativi di  $T$  è trascurabile quando  $\theta = \delta > 0$  perciò per calcolare la potenza si usa

$$\Pr\{T(mK - p; \delta / \sqrt{R s^2 c^T (D^{(k)T} D^{(k)})^{-1} c}) \geq t_{mK-p, 1-\alpha/2}\}. \quad (2.15)$$

La potenza per  $\theta = -\delta$  è uguale, data la simmetria.

Una forma più grezza ma più semplice per calcolare il livello di potenza che un test  $t$  sequenziale raggiunge è

$$\Phi\left(\frac{\delta}{\sqrt{R s^2 c^T (D^{(k)T} D^{(k)})^{-1} c}} - \Phi^{-1}(1 - \alpha/2)\right) \quad (2.16)$$

che si utilizza quando  $mK - p$  è grande e si sceglie di trascurare l'errore derivato dalla stima di  $\sigma^2$ .

In Shao e Feng (2007) viene considerato il test  $t$  sequenziale per valori piccoli di  $m$ . I valori delle soglie dovrebbero essere calcolati in riferimento alla probabilità (2.14), perciò si necessita della distribuzione congiunta di

$(T_1, \dots, T_K)$ . Dato che tale distribuzione non ha una forma semplice si è scelto di derivarla tramite simulazione utilizzando il metodo Monte Carlo. La probabilità di errore di I tipo da soddisfare è data dalla (2.14) e tramite simulazioni sono stati ottenuti i valori di  $(c_1, \dots, c_K)$  che soddisfano l'uguaglianza in base al valore di  $\alpha$  scelto. Shao e Feng (2007) dimostrano che tali soglie sono più accurate nel rispettare le probabilità di errore di I e II tipo.

## 2.8 Test unilaterali

Quando l'ipotesi nulla è contrapposta ad un'alternativa unilaterale, per esempio  $H_0 : \theta = 0$  contro  $H_1 : \theta > 0$  si parla di problemi di verifica d'ipotesi con alternativa unilaterale. In molti casi è appropriato considerare problemi di questo tipo, per esempio quando verificare l'ipotesi alternativa in una delle due direzioni è insensato o impossibile o quando sono di interesse solo variazioni in una certa direzione.

A differenza dei corrispondenti test bilaterali, i test unilaterali sono definiti da coppie di costanti  $(a_k, b_k)$  con  $a_k < b_k$  per  $k = 1, \dots, K-1$ , e  $a_K = b_K$ . Si supponga di utilizzare la statistica test  $Z_{ep,k}$ . La regola d'arresto è definita nel seguente modo: alla  $k$ -esima analisi, per  $k = 1, \dots, K-1$ , se  $Z_{ep,k} \geq b_k$  l'analisi si conclude rifiutando  $H_0$ , se  $Z_{ep,k} \leq a_k$  l'analisi si conclude accettando  $H_0$ , altrimenti continua analizzando il gruppo successivo. Alla  $K$ -esima analisi, se  $Z_{ep,K} \geq b_K$  si rifiuta  $H_0$  mentre se  $Z_{ep,K} < a_K$  si conclude accettando la nulla. Definire  $a_K = b_K$  assicura che il test termini alla  $K$ -esima analisi in favore di una delle due ipotesi.

Nell'ambito di questo tipo di test, è interessante ricordare la proposta della famiglia di test unilaterali fatta da Emerson & Fleming (1989) indicizzata ad un parametro  $\Delta$ . Più  $\Delta$  cresce maggiori sono le possibilità di arresto anticipato, e maggiore diventa la numerosità campionaria massima, poiché il fattore di proporzionalità con un test a numerosità campionaria fissata,  $\tilde{R}(K, \alpha, \beta, \Delta)$ , che verrà introdotto a breve, cresce con  $\Delta$ . Si utilizza la successione di statistiche test  $Z_{ep,k}$  con parametro di interesse  $\theta$  di dimensione

1 con distribuzione congiunta (2.9) uguale a quella dei corrispondenti test bilaterali, dato che l'ipotesi nulla di uguaglianza a zero è la stessa. Il test è condotto ipotizzando equispaziatura tra i livelli di informazione. I valori critici per  $k = 1, \dots, K$  sono

$$\begin{aligned} b_k &= \tilde{C}_1(K, \alpha, \beta, \Delta)(k/K)^{\Delta-1/2} \\ a_k &= \delta\sqrt{V_k(\theta)} - \tilde{C}_2(K, \alpha, \beta, \Delta)(k/K)^{\Delta-1/2}. \end{aligned} \quad (2.17)$$

Per fare in modo che  $a_K = b_K$  l'informazione finale deve essere

$$V_K(\theta) = \frac{\{\tilde{C}_1(K, \alpha, \beta, \Delta) + \tilde{C}_2(K, \alpha, \beta, \Delta)\}^2}{\delta^2}. \quad (2.18)$$

Le costanti  $\tilde{C}_1(K, \alpha, \beta, \Delta)$  e  $\tilde{C}_2(K, \alpha, \beta, \Delta)$ , indipendenti da  $\theta$ , sono scelte in modo da garantire i valori obiettivo degli errori di I e II tipo. In test di questo tipo, il vincolo superiore può essere visto come un test di significatività ripetuta per l'ipotesi  $\theta = 0$  con valori critici per  $Z_{ep,k}$  proporzionali a  $k^{\Delta-1/2}$  mentre la soglia inferiore si riconduce ad un test ripetuto con ipotesi nulla  $\theta = \delta$  anch'essa con valori critici per  $Z_k - \delta\sqrt{V_k(\theta)}$  proporzionali a  $k^{\Delta-1/2}$ .

Utilizzando la (2.18) ed esprimendo la (2.3) in termini di informazione necessaria per un test a numerosità campionaria fissata unilaterale, si ottiene

$$\tilde{R}(K, \alpha, \beta, \Delta) = \frac{V_K(\theta)}{V(\theta)} = \frac{\{\tilde{C}_1(K, \alpha, \beta, \Delta) + \tilde{C}_2(K, \alpha, \beta, \Delta)\}^2}{\{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta)\}^2 / \delta^2}.$$

che è il rapporto tra l'informazione massima necessaria per un test sequenziale a gruppi rispetto a quella richiesta per un test a numerosità fissa. Il fattore  $\tilde{R}(K, \alpha, \beta, \Delta)$  è utilizzato per calcolare il livello di informazione massimo per soddisfare i requisiti di potenza. Valori tabulati di  $\tilde{R}(K, \alpha, \beta, \Delta)$ ,  $\tilde{C}_1(K, \alpha, \beta, \Delta)$  e  $\tilde{C}_2(K, \alpha, \beta, \Delta)$  per diversi  $K, \alpha, \beta, \Delta$  si possono trovare in Emerson e Fleming (1989) o Pampallona e Tsiatis (1994).

## 2.9 Alcuni test sequenziali

Di seguito verranno introdotti brevemente i test di Pocock (1977), O'Brien e Fleming (1979) e Wang e Tsiatis (1987), basandosi sul volume di Jennison e Turnbull (2000, Capitolo 2), al quale si rimanda per approfondimenti. Si consideri un test sequenziale comparativo, con  $K$  analisi, in cui si osservano risposte da soggetti assegnati a due diversi trattamenti, uno sperimentale, E, e uno di controllo, C. Le risposte sono assunte normali indipendenti con varianza nota,  $X_i \sim N(\mu_E, \sigma^2)$  e  $Y_i \sim N(\mu_C, \sigma^2)$ , dove  $i = 1, 2, \dots, mk$ . Il parametro di interesse è la differenza tra le medie dei due gruppi:  $\theta = \mu_E - \mu_C$  e si desidera verificare l'ipotesi nulla  $H_0 : \theta = 0$  contro alternativa bilaterale. Si ha che  $\text{Var}(\hat{\theta})$  coincide con il reciproco dell'informazione attesa parziale per  $\theta$ ,  $V_k(\theta)^{-1}$ . La statistica utilizzata per verificare le ipotesi è  $Z_{ep,k}$ . Viene calcolata ad ogni analisi e ha la forma

$$Z_{ep,k} = \frac{1}{\sqrt{(2mk\sigma^2)}} \left( \sum_{i=1}^{mk} X_i - \sum_{i=1}^{mk} Y_i \right), \quad k = 1, \dots, K. \quad (2.19)$$

che avrà distribuzione  $N((\mu_E - \mu_C)\sqrt{mk/2\sigma^2}, 1)$ , cioè  $N(\theta\sqrt{V_k(\theta)}, 1)$ . Questo è il modello di base per i test descritti nel seguito.

### 2.9.1 Il test di Pocock

Nel test di Pocock la regione di arresto è definita tramite le costanti  $C_P(K, \alpha)$ . Tali valori sono scelti in modo da mantenere la probabilità di errore di I tipo uguale ad  $\alpha$  per l'esperimento nel suo complesso, distribuendola in parti uguali ad ogni analisi. Le soglie devono soddisfare l'uguaglianza

$$\Pr_{\theta=0} \{ \text{Rifiutare } H_0 \text{ all'analisi } k = 1, k = 2, \dots, \text{ o } k = K \} = \alpha.$$

Formalmente la regola d'arresto funziona come segue: dal primo al  $(K - 1)$ -esimo gruppo l'esperimento termina se si rifiuta  $H_0$ , cioè se  $|Z_{ep,k}| \geq C_P(K, \alpha)$ , altrimenti continua passando al gruppo successivo. Se l'esperime-

to prosegue fino all'ultimo gruppo, il  $K$ -esimo, si sceglie in favore dell'ipotesi alternativa se  $|Z_{ep,K}| \geq C_P(K, \alpha)$  e in favore di  $H_0$  altrimenti.

Le soglie  $C_P(K, \alpha)$  sono calcolate numericamente utilizzando la distribuzione congiunta della successione di statistiche  $(Z_{ep,1}, \dots, Z_{ep,K})$ . Ad esempio, per  $K = 2$ , il valore  $C_P(2, \alpha)$  deriva dai seguenti passaggi:

$$\begin{aligned}
& P(\text{Rifiutare } H_0 | \theta = 0) = \\
& = P\{(|Z_{ep,1}| \geq C_P(2, \alpha)) \text{ o } (|Z_{ep,1}| \leq C_P(2, \alpha) \text{ e } Z_{ep,2} \geq C_P(2, \alpha)) \text{ o} \\
& \quad (|Z_{ep,1}| \leq C_P(2, \alpha) \text{ e } Z_{ep,2} \leq -C_P(2, \alpha))\} \\
& = P\{|Z_{ep,1}| \geq C_P(2, \alpha)\} + 2P\{Z_{ep,1} \geq -C_P(2, \alpha) \text{ e } Z_{ep,2} \geq C_P(2, \alpha)\} \\
& \quad - 2P\{Z_{ep,1} \geq C_P(2, \alpha) \text{ e } Z_{ep,2} \geq C_P(2, \alpha)\} \\
& = P\{|Z_{ep,1}| \geq C_P(2, \alpha)\} + 2P\{-Z_{ep,1} \leq C_P(2, \alpha) \text{ e } Z_{ep,2} \leq -C_P(2, \alpha)\} \\
& \quad - 2P\{-Z_{ep,1} \leq -C_P(2, \alpha) \text{ e } -Z_{ep,2} \leq -C_P(2, \alpha)\} \\
& = 2[1 - \Phi(C_P(2, \alpha))] + \Phi_2(C_P(2, \alpha), -C_P(2, \alpha), \rho) \\
& \quad - \Phi_2(-C_P(2, \alpha), -C_P(2, \alpha), \rho)
\end{aligned} \tag{2.20}$$

dove  $\Phi_2(x_1, x_2, \rho) = P(X_1 \leq x_1, X_2 \leq x_2)$  per  $X = (X_1, X_2)^T \sim N_2(0, \Sigma)$  con  $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ . Il valore di  $C_P(2, \alpha)$  è scelto in modo da rendere i passaggi sopra uguali ad  $\alpha$ .

Nel test di Pocock ad ogni analisi la probabilità di errore di I tipo è sempre uguale per ogni analisi ed è  $\alpha' = 2[1 - \Phi\{C_P(K, \alpha)\}]$ .

Perciò il test rifiuta  $H_0$  se il livello di significatività osservato  $\alpha^{oss}$  per un test non sequenziale su  $H_0$  con alternativa bilaterale che utilizza i dati disponibili fino alla  $k$ -esima analisi è minore di  $\alpha'$ . La probabilità di errore di primo tipo si mantiene pari ad  $\alpha$  qualunque sia la numerosità dei gruppi  $m$ .

I requisiti di potenza sono

$$\Pr_{\theta=\pm\delta} \{\text{Rifiutare } H_0\} = 1 - \beta.$$

e vengono soddisfatti scegliendo il livello di informazione massimo,  $V_{max}(\theta) = V_K(\theta)$ , che verifica questa uguaglianza. Essendo in questo caso  $V_K(\theta) = mK/2\sigma^2$ , con  $\sigma^2$  noto, questo si traduce in una scelta del valore di  $mK$ . Arrotondamenti di  $m$  ad un valore intero possono avere un leggero effetto sulla potenza del test, perciò è consigliabile arrotondarli in eccesso, in modo da ottenere un test conservativo, cioè con potenza di poco superiore a quella nominale.

Come visto per i test sequenziali in generale, anche per quelli di Pocock il livello di informazione massima richiesto è proporzionale a quella di un test a numerosità campionaria fissa. La costante di proporzionalità è data da:

$$R_P(K, \alpha, \beta) = \frac{V_K(\theta)}{V(\theta)}$$

e i valori sono raccolti in tabelle per varie combinazioni di  $K$ ,  $\alpha$  e  $\beta$ . Si veda, ad esempio, Jennison e Turnbull (2000, Capitolo 2). Le costanti di proporzionalità vengono utilizzate per definire l'informazione massima necessaria per il test sequenziale, a partire da quella di un test a numerosità campionaria fissata.

I valori di  $R_P(K, \alpha, \beta)$  derivano dallo stesso tipo di calcoli numerici utilizzati per determinare  $C_P(K, \alpha)$ .

### 2.9.2 Il test di O'Brien e Fleming

In alternativa al test appena visto, O'Brien e Fleming (1979) propongono un test in cui il livello di significatività nominale, non viene mantenuto uguale per tutte le analisi, ma cresce al procedere dello studio. Di conseguenza risulterà più difficile rifiutare l'ipotesi nulla nelle prime analisi ma più facile successivamente.

Il test funziona come quello di Pocock, ma cambia la successione di valori critici  $c_1, \dots, c_K$  che in questo caso derivano dalla moltiplicazione di una costante  $C_B(K, \alpha)$  per  $\sqrt{K/k}$ . In tal modo le soglie  $c_k$  decrescono al crescere di  $k$ , ed il livello nominale di significatività cresce. Come nel test di

Pocock, le costanti  $C_B(K, \alpha)$  sono calcolate in modo da garantire la probabilità complessiva di errore di I tipo pari ad  $\alpha$ . Si noti che quando  $K = 1$ ,  $C_B(1, \alpha) = C_P(1, \alpha)$ . Cambiando le barriere, cambiano anche le costanti di proporzionalità per il calcolo della numerosità campionaria massima. Tali costanti sono riportate in tabelle simili a quelle per il test di Pocock e sono calcolate allo stesso modo. Si veda, ad esempio, Jennison e Turnbull (2000, Capitolo 2).

### 2.9.3 I test di Wang e Tsiatis

Questa famiglia di test bilaterali, proposta da Wang e Tsiatis (1987) è indicizzata da un parametro  $\Delta$  al variare del quale variano le forme delle barriere, ed include i test di Pocock e O'Brien e Fleming come casi particolari. Il test rifiuta  $H_0$  dopo l'analisi del gruppo  $k$  se  $|Z_{ep,k}| \geq c_k$  dove  $c_k = C_{WT}(K, \alpha, \Delta)(k/K)^{\Delta-1/2}$ ,  $k = 1, \dots, K$ , altrimenti prosegue con l'analisi del gruppo successivo. Se si arriva all'ultimo gruppo,  $K$ , si rifiuta  $H_0$  se  $|Z_k| \geq C_{WT}(K, \alpha, \Delta)$  altrimenti si accetta  $H_0$ .

Prendendo  $\Delta = 0.5$  o  $\Delta = 0$  si ottengono i test di Pocock e O'Brien e Fleming rispettivamente. Valori di  $\Delta$  tra 0 e 0.5 restituiscono test con barriere di forma intermedia. Come per i precedenti test, le costanti  $C_{WT}(K, \alpha, \Delta)$  sono scelte per assicurare un errore di I tipo globale pari ad  $\alpha$ . Per la numerosità campionaria, calcoli numerici hanno fornito i valori  $R_{WT}(K, \alpha, \beta, \Delta)$  aventi la stessa funzione di  $R_P(K, \alpha, \beta)$  e  $R_B(K, \alpha, \beta)$  viste nei precedenti test.

I test di Wang e Tsiatis sono una famiglia molto ricca e flessibile. Di particolare interesse sono quelli con  $\Delta = 0.25$  caratterizzati da ampie soglie nelle prime analisi, che rendono più difficile l'arresto anticipato, ma non troppo estreme come quelle dei test di O'Brien e Fleming. Questa famiglia di test gode di alcuni dei vantaggi dei test di Pocock e O'Brien e Fleming. Rispetto al primo ha in comune bassa numerosità campionaria attesa quando il vero valore del parametro di interesse è distante dall'ipotesi nulla. Rispetto al test di O'Brien e Fleming ha in comune bassa numerosità campionaria massima

e valori critici vicini a quelli di un test a campione fisso al raggiungimento dell'ultima analisi.

### Esempio: Test di Wang e Tsiatis con risposte normali

A titolo illustrativo si riporta ora un esempio del test di Wang e Tsiatis tratto da Jennison e Turnbull (2000, Paragrafo 2.7.1). È possibile derivare un esempio per il test di Pocock semplicemente modificano il valore di  $\Delta$ , ponendolo pari a 0.5. Lo stesso si può fare per il test di O'Brien e Fleming, definendo  $\Delta = 0$ .

Si ipotizzi di condurre un esperimento sequenziale, con  $K = 5$  gruppi di osservazioni,  $\sigma^2 = 4$ ,  $\alpha = 0.05$  e potenza  $1 - \beta = 0.9$  per  $\mu_E - \mu_C = \pm 1$ . Il parametro di interesse è al solito  $\theta = \mu_E - \mu_C$  che vale 0 sotto  $H_0$ . La varianza dello stimatore di massima verosimiglianza è il reciproco dell'informazione attesa parziale,  $V_k(\theta)$ . L'esperimento viene condotto utilizzando le barriere di Wang e Tsiatis con  $\Delta = 0.25$  e ad ogni analisi viene calcolata la statistica test  $Z_{ep,k}$ .

La numerosità campionaria necessaria per garantire i requisiti di potenza viene calcolata utilizzando l'equazione  $V_K(\theta) = R_B(K, \alpha, \beta)V(\theta)$ , con  $V(\theta) = n_{f,2}/(2\sigma^2)$  e  $V_K(\theta)$  con formulazione analoga ma calcolata su  $mK$  osservazioni. La numerosità necessaria per un test a campione fisso con alternativa bilaterale, utilizzando la (2.2) è  $n_{f,2} = 84.1$  osservazioni per tipo di trattamento. Il valore della costante di proporzionalità, è  $R_{WT}(K, \alpha, \beta, \Delta) = 1.066$  (Tabella 2.10, Jennison e Turnbull, 2000). Si ottiene quindi  $mK = 1.066 \times 84.1$ . Con  $K = 5$  i requisiti di potenza sono soddisfatti se le analisi intervallari sono svolte ogni  $m = 17.93$  pazienti per tipo di trattamento, arrotondato a 18.

Ad ogni analisi viene ricalcolata la statistica test aggiungendo le nuove risposte osservate. Le soglie hanno la forma  $C_{WT}(K, \alpha, \Delta)(k/K)^{\Delta-1/2}$  e in questo caso  $C_{WT}(5, 0.05, 0.25) = 2.136$  (Tabella 2.9, Jennison e Turnbull,



2000). Il test termina rifiutando l'ipotesi nulla alla  $k$ -esima analisi se:

$$|Z_{ep,k}| \geq C_W T(K, \alpha, \Delta) (k/K)^{\Delta-1/2},$$

cioè

$$\frac{1}{\sqrt{18k \times 2 \times 4}} \left| \sum_{i=1}^{18k} X_i - \sum_{i=1}^{18k} Y_i \right| \geq 2.136 (k/5)^{-0.25}$$

e se questo non avviene per nessun  $k = 1, \dots, 5$ , allora si accetta  $H_0$ . Il numero massimo di risposte osservate per l'esperimento nel complesso è  $2mk = 180$ .

#### 2.9.4 Proprietà e confronto

Quando si confrontano test sequenziali, si può decidere quali proprietà considerare. Normalmente le più indicative sono le probabilità di errore di I e II tipo, la numerosità campionaria attesa e quella massima. Quest'ultima indica cosa potrebbe succedere nel caso peggiore. Nella Tabella 2.1, tratta da Jennison e Turnbull (2000, Tabella 2.5), è riportato un esempio di confronto tra il test di Pocock e O'Brien & Fleming con  $K=5$ . Ciascuno dei test considerati finora ha vantaggi e svantaggi. Il test di Pocock ha numerosità campionaria attesa minore quando  $|\theta|$  è grande; perciò asseconda in misura maggiore i requisiti etici. Tuttavia quando  $|\theta|$  è piccolo, la numerosità campionaria attesa è più alta e addirittura superiore a quella di un'analisi a campione fisso.

Il test di O'Brien e Fleming, avendo il livello di significatività nominale dell'ultima analisi vicino alla probabilità di errore di I tipo totale, tutela contro la situazione ambigua in cui l'ipotesi nulla è accettata mentre verrebbe rifiutata in un'analisi a campione fisso che consideri tutti i dati raccolti.

In alcuni casi, per varie ragioni, può essere necessario scegliere un test che renda difficile l'arresto durante le prime analisi. Per esempio, problemi tecnici nella qualità dei dati, risolti poco dopo l'inizio dell'analisi o difficoltà nel controllare le assunzioni fatte sulla distribuzione dei dati per un campione

**Tabella 2.1:** Funzione di potenza, numerosità campionaria attesa e sua deviazione standard per tre tipi di test. I test hanno 5 gruppi di osservazioni, errore di primo tipo  $\alpha = 0.05$  e potenza  $1 - \beta = 0.9$  per  $|\mu_A - \mu_B| = 1$ . Le osservazioni per ogni trattamento sono normali con varianza  $\sigma^2 = 4$ . Fonte: Jennison e Turnbull (2000, Tabella 2.5).

	Campione Fisso	Pocock	O'Brien e Fleming
		Pr(Rifiutare $H_0$ )	
$ \mu_A - \mu_B  = 0.0$	0.050	0.050	0.050
0.5	0.371	0.351	0.378
1.0	0.903	0.910	0.912
1.5	0.998	0.999	0.999
		E(Nr oss)	
$ \mu_A - \mu_B  = 0.0$	170.0	204.8	178.7
0.5	170.0	182.3	167.9
1.0	170.0	116.9	129.8
1.5	170.0	70.1	94.4
		Dev. Std. (Nr oss)	
$ \mu_A - \mu_B  = 0.0$	0.0	26.1	8.6
0.5	0.0	50.8	24.7
1.0	0.0	57.9	35.5
1.5	0.0	34.1	25.7

molto piccolo. Di grande rilevanza anche la consapevolezza che la comunità scientifica potrebbe non accettare conclusioni basate solo su poche osservazioni. Tali questioni sono di minore interesse se la numerosità campionaria è sostanziale già dalla prima analisi; allora la commissione di controllo dell'esperimento potrebbe essere riluttante a continuare un esperimento dopo aver osservato un valore del test molto alto nelle prime analisi, anche se il vincolo di O'Brien & Fleming non è stato superato. Non sempre perciò la scelta è semplice. Il test di Wang & Tsiatis con  $\Delta = 0.25$ , che gode di molte delle proprietà di quelli di Pocock e O'Brien & Fleming, può essere visto come un compromesso tra i due.

# Capitolo 3

## Il test triangolare di Whitehead

### 3.1 Introduzione

Un test sequenziale a gruppi alternativo a quelli descritti nel Capitolo 2, è quello triangolare proposto da Whitehead. Ad esso verrà dedicato il presente capitolo, trattandolo più approfonditamente degli altri test, in quanto il lavoro svolto per questa tesi è basato su di esso. Si farà principalmente riferimento a quanto riportato nel libro di Whitehead (1997). Come i test visti in precedenza, anche il test triangolare si può descrivere basandosi su quantità di verosimiglianza ed utilizza opportune barriere per verificare l'ipotesi di interesse. A differenza degli altri test ha il vantaggio di poter essere applicato senza prefissare il numero massimo di analisi,  $K$ , o il livello di informazione necessario per ogni analisi e soddisfa accuratamente i requisiti delle probabilità di errore di I e II tipo anche quando le analisi non sono tra loro equispaziate nei livelli di informazione. Inoltre, per determinare le barriere sono sufficienti semplici calcoli, poichè vengono approssimate a delle rette.

Il Paragrafo 2 descrive le caratteristiche principali del test triangolare. Il Paragrafo 3 estende quanto esposto nel Paragrafo 2 nei casi in cui l'ipotesi alternativa da verificare sia bilaterale o unilaterale. Il Paragrafo 4 riassume gli strumenti utilizzati per analizzare l'esperimento una volta concluso.

## 3.2 Caratteristiche principali dell'esperimento

Come fatto in precedenza, ci si concentrerà su esperimenti comparativi, tra due trattamenti, E e C. Il parametro di interesse  $\theta$  è indice della differenza media tra due trattamenti ed è scalare, mentre il parametro di disturbo  $\lambda$  ha dimensione  $p - 1$ . Si è interessati a verificare l'ipotesi nulla di uguaglianza dei due trattamenti. Se l'ipotesi alternativa è unilaterale, si utilizza il test triangolare singolo, mentre nel caso di ipotesi alternativa bilaterale, si utilizza il test triangolare doppio.

Il test si basa principalmente su due quantità, la *score*,  $S_k(0)$ , e l'informazione attesa parziale,  $V_k(0)$ , entrambi calcolati per  $\theta = 0$ . Da questo momento, per alleggerire la notazione,  $S_k(0)$  e  $V_k(0)$  verranno indicati con  $S_k$  e  $V_k$ . Si ricorda la forma di  $S_k$  e  $V_k$  in termini di quantità di verosimiglianza:

$$S_k = l_\theta(0, \hat{\lambda}_0^{(k)}) \quad \text{e} \quad V_k = \{i^{\theta\theta}(0, \hat{\lambda}_0^{(k)})\}^{-1},$$

dove  $\hat{\lambda}_0^{(k)}$  è la stima di massima verosimiglianza di  $\lambda$  per  $\theta$  fissato pari a 0. In generale,  $S_k(\theta)$  ha distribuzione asintotica normale di media  $\theta V_k(\theta)$  e varianza  $V_k(\theta)$ , e per  $\theta = 0$ ,  $S_k/\sqrt{V_k}$  ha distribuzione approssimata normale standard (tale distribuzione è esatta nel caso di osservazioni normali con varianza nota).

Essenzialmente, il test triangolare di Whitehead utilizza il test *score*,  $Z_{up}$ , discusso nel Paragrafo 2.2.1. Nel Capitolo 4 verrà considerato, in alternativa, il test  $Z_p$  e un suo miglioramento basato su argomenti asintotici di ordine superiore.

Il test triangolare è sviluppato approssimando il processo *score* con un moto Browniano con deriva (*drift*)  $\theta$  e varianza unitaria, osservato ai tempi  $V_1, \dots, V_K$ . La teoria sviluppata da Whitehead (1997, Paragrafi 4.11-4.14) consente di ottenere espressioni semplici per barriere lineari del tipo

$$\begin{aligned} u_k &= a + cV_k \\ l_k &= -a + 3cV_k, \end{aligned}$$

a partire da approssimazioni per la funzione di potenza ottenute ignorando l'effetto (*overshoot*) dovuto al monitoraggio non continuo. In particolare, assumendo che la distribuzione congiunta delle statistiche  $S_k$  sia quella di un moto Browniano  $\{S(V), V > 0\}$  il test triangolare considera barriere lineari sul piano  $(V, S)$ .

La regola di arresto è applicata come segue. Siano  $u_k$  e  $l_k$  i valori che assumono la barriera superiore e inferiore rispettivamente all'analisi  $k$ , cioè in corrispondenza di  $V_k$ . Se  $S_k \geq u_k$  il test termina rifiutando  $H_0$ , indicando una superiorità del trattamento alternativo. Se  $S_k \leq l_k$  il test termina accettando  $H_0$ . Negli altri casi non si ha nessuna evidenza e si prosegue all'analisi successiva. Arrivati all'ultima analisi, il test termina rifiutando  $H_0$  se  $S_K \geq u_K$ , accettandola altrimenti.

Per calcolare i valori critici si utilizzano i requisiti di potenza da soddisfare, basati sulle probabilità di errore di I e II tipo,  $\alpha$  e  $\beta$ . In base a quanto riportato da Whitehead (1997) si adotta un approccio simmetrico. Cioè, definendo  $C(0) = C^+(0) + C^-(0) = \alpha$ , si impone che  $C^+(0) = C^-(0) = \alpha/2$ , dove  $C^+(0)$  è la probabilità sotto  $H_0$  di rifiutare l'ipotesi nulla a favore dell'ipotesi di superiorità del trattamento sperimentale, e viceversa per  $C^-(0)$ . Si definisce inoltre  $C(\delta) = C^+(\delta) + C^-(\delta)$ , con  $C^+(\delta)$  che è la probabilità di rifiutare l'ipotesi nulla a favore dell'ipotesi di superiorità del trattamento sperimentale quando  $\theta = \delta$ , e viceversa per  $C^-(\delta)$ . Il valore  $\delta$  rappresenta il valore del parametro sotto l'ipotesi alternativa. Quando si desidera verificare un'ipotesi con alternativa  $H_1 : \theta > 0$ , si utilizza il test triangolare singolo.

Di seguito si farà inizialmente riferimento al test triangolare singolo, poiché il test triangolare doppio rappresenta una semplice estensione e utilizza le stesse quantità utilizzate per il caso di alternativa unilaterale. Si è interessati quindi a soddisfare i requisiti  $C^+(0) = \alpha/2$  e  $C^+(\delta) + C^-(\delta) = 1 - \beta$ . Dato che per ipotesi di questo tipo il valore di  $C^-(\delta)$  è tipicamente molto piccolo, l'ultima uguaglianza diventa  $C^+(\delta) = 1 - \beta$ ; non si ha nessun vincolo sul valore di  $C^-(\delta)$  e  $C^-(0)$ .

In un test triangolare singolo, in cui si desidera verificare se i due trat-

tamenti sono uguali, contro l'alternativa che il trattamento sperimentale sia superiore a quello di controllo (indicato con  $E \succ C$ ), i requisiti di potenza sono della forma:

$$\begin{aligned} C(0) &= P(\text{concludere } E \succ C | \theta = 0) = \alpha/2 \\ C(\delta) &= P(\text{concludere } E \succ C | \theta = \delta) = 1 - \beta \end{aligned}$$

con  $\delta > 0$ .

In un test a campione fisso con  $\alpha$  e  $\beta$  fissati, il valore critico che soddisfa il primo requisito, è  $c\sqrt{V_k}$ , con  $c = \Phi^{-1}(1 - \alpha/2)$ , essendo che sotto l'ipotesi nulla  $S \sim N(0, V)$ . Una volta trovato  $c$ , utilizzando l'equazione di  $C(\delta)$ , è possibile definire l'informazione  $V$  necessaria a soddisfare i requisiti di potenza. Sapendo che per  $\theta = \delta$ ,  $S(\delta) \sim N(\delta V(\delta), V(\delta))$ , si ottiene  $V(\delta) = \{c + \Phi^{-1}(1 - \beta)\}^2 / \delta^2$ .

Più complessa è la situazione per un test triangolare. Si ipotizzi, per semplicità, di effettuare  $K = 2$  analisi. La prima viene condotta calcolando  $V_1$ ,  $l_1$ ,  $u_1$  e  $S_1$ . Se  $S_1 \geq u_1$  lo studio si arresta concludendo che  $E \succ C$ . Se  $S_1 \leq l_1$  lo studio si ferma concludendo per l'uguaglianza dei due trattamenti. Se  $l_1 < S_1 < u_1$  si procede all'analisi successiva e vengono calcolate le stesse quantità per  $k = 2$ . A questo punto se  $S_2 \geq u_2$  si conclude che  $E \succ C$ , altrimenti si conclude che non c'è differenza tra i due trattamenti.

Rispetto ai test visti nel Capitolo 2, il calcolo delle costanti da utilizzare per le barriere risulta un po' più complesso. Ad esempio, riferendosi ai calcoli esatti visti nelle (2.20), nel test triangolare le incognite sarebbero  $l_1, u_1, u_2, V_1, V_2$  che devono soddisfare le due equazioni dei requisiti di potenza. Per poter definire univocamente ogni variabile, si dovrebbero definire ulteriori vincoli.

Nel caso in cui  $K > 2$ , è evidente come diventa ancora più laborioso effettuare considerazioni individuali su ogni valore soglia. Al fine di evitare calcoli troppo complessi, Whitehead (1997, Paragrafo 4.11) propone di utilizzare valori soglia lineari. In particolare propone di approssimarli con delle

rette di equazione

$$\begin{aligned}u_k &= a + cV_k \\l_k &= -a + 3cV_k,\end{aligned}$$

che definiscono la regione di continuazione del test. I valori di  $a$  e  $c$  sono calcolati in modo da soddisfare i requisiti di potenza. Espressioni semplici esistono solo quando  $\delta$  è un multiplo di  $2c$  e quando  $\frac{1}{2}\alpha = \beta$ . Nel seguito ci si concentrerà solamente sul caso in cui  $\theta = 0$  sotto l'ipotesi nulla e  $\theta = \delta = 4c$  per fissare la potenza sotto l'ipotesi alternativa; in questa fattispecie si ha

$$\begin{aligned}a &= (2/\delta) \log(1/\alpha), \\c &= \frac{1}{4}\delta.\end{aligned}\tag{3.1}$$

Nel seguito verrà sempre utilizzata questa formulazione per le barriere, perciò si avrà in ogni esempio  $\frac{1}{2}\alpha = \beta$ .

Per completezza, si noti che, nel caso in cui  $\frac{\alpha}{2} \neq \beta$ , in un test a campione fisso con errore di I tipo  $\alpha$  e potenza  $1 - \beta$  in  $\theta = \delta$ , la potenza è pari a  $1 - \alpha$  per  $\theta = \xi\delta$ , dove

$$\xi = \frac{2\Phi^{-1}(1 - \alpha)}{\Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta)}.$$

Il test può quindi essere condotto come definito sopra ma utilizzando  $\tilde{\delta} = \xi\delta$  al posto di  $\delta$  nelle formule per il calcolo delle barriere e di  $V$  (Whitehead e Stratton, 1983). Non verranno trattati qui esempi di questo tipo. Per collegare il questo tipo di barriere a quelle dei test visti nel Capitolo 2, si veda Kittelson e Emerson (1999).

In fase di progettazione di un test triangolare, non è necessario definire i livelli di informazione da raggiungere ad ogni analisi, ma è importante definire qual'è l'informazione che consente di soddisfare i requisiti di potenza, ossia l'informazione massima,  $V_{max}$ . Come si vede dalla Figura 3.1 le due barriere delimitano un'area di forma triangolare, chiamata **regione di continuazione**. Le due soglie si avvicinano al crescere di  $V_k$ ,

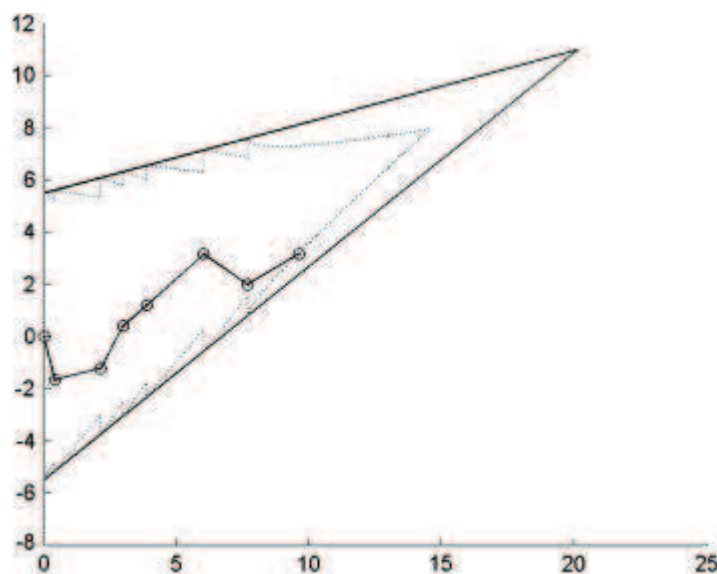
fino ad incontrarsi nel punto  $V_{max} = a/c$ . Il vertice del triangolo rappresenta un limite superiore per l'informazione attesa, e dato che le barriere sono calcolate tenendo conto dei requisiti di potenza, consente di trovare  $V_{max}$ . Al valore di  $V_{max}$  corrisponde una certa numerosità campionaria, che è la numerosità campionaria massima dell'intero esperimento,  $mK$ .

È importante notare che  $V_k$  viene utilizzato solo in fase di progettazione dell'esperimento, poiché è necessario per definire  $mK$ . Una volta iniziato l'esperimento  $V_k$  è sostituito da  $I_k(0)$  l'informazione osservata parziale calcolata utilizzando i dati disponibili alla  $k$ -esima analisi. Per brevità, verrà indicata con  $I_k$ . Quando un esperimento viene condotto fino all'ultima analisi, e il valore di  $V_{max}$  non dipende dalle osservazioni, dovrebbe verificarsi che  $I_K = V_{max}$ . Se  $V_{max}$  dipende dalle osservazioni è possibile riscontrare qualche differenza tra l'informazione attesa massima e l'informazione osservata finale, in seguito alla variabilità dei dati.

Bisogna considerare che il monitoraggio continuo non è ottenibile in pratica; anche analizzando i dati dopo ogni singola risposta non si impedisce a  $V_k$  di crescere a piccoli salti. Inoltre svolgere le indagini a determinati istanti di tempo rende difficile controllare i valori della statistica test nell'intervallo tra un'analisi e l'altra. Per applicare quindi vincoli continui ad esperimenti che di fatto sono discreti, Whitehead e Stratton (1983) hanno adattato i test continui ad un monitoraggio discreto, utilizzando un risultato di Siegmund (1979). Tali aggiustamenti, come si può vedere in Figura 3.1 vengono chiamati ad **albero di Natale** per via della loro forma.

Sostanzialmente le soglie del test originale vengono portate verso l'interno per un ammontare pari a  $0.583\sqrt{(V_k - V_{k-1})}$  ad ogni analisi. Quindi più è elevata la differenza tra le informazioni di due analisi successive, maggiore sarà l'aggiustamento necessario. Perciò, per una generica successione di





**Figura 3.1:** Test triangolare singolo disegnato nel piano  $(V, S)$ . Le due linee continue rappresentano le due barriere, e quelle tratteggiate rappresentano la correzione ad albero di Natale. All'interno è disegnato un esempio di *sample path* ossia di valori che può assumere la statistica  $S_k$  ad ogni analisi. Fonte: <http://www.sciencedirect.com>.

$(V_1, \dots, V_K)$  le barriere diventano

$$u_k = \frac{2}{\delta} \log\left(\frac{1}{2\alpha}\right) - 0.583\sqrt{V_k - V_{k-1}} + \frac{\delta}{4}V_k,$$

$$l_k = -\frac{2}{\delta} \log\left(\frac{1}{2\alpha}\right) + 0.583\sqrt{V_k - V_{k-1}} + \frac{3\delta}{4}V_k.$$

La costante 0.583 è una correzione di continuità e deriva da calcoli approssimati sviluppati in Siegmund (1985). Nel caso in cui si utilizzi la correzione di continuità, le soglie vengono avvicinate tra loro, e il valore massimo di  $V_K$  diminuisce. Se si considerano analisi effettuate a livelli di informazioni equidistanti, con  $V_K = V_{max}$ , l'intervallo tra le informazioni è pari a  $V_{max}/K$ , e  $V_k = (k/K)V_{max}$ . Il livello di informazione massimo,  $V_{max}$ , utilizzando la

correzione di continuità verrà calcolato come punto di incontro tra le rette

$$\begin{aligned} u_k &= a + c \frac{k}{K} V_{max} - 0.583 \sqrt{\frac{V_{max}}{K}} \\ l_k &= -a + 3c \frac{k}{K} V_{max} + 0.583 \sqrt{\frac{V_{max}}{K}}. \end{aligned} \quad (3.2)$$

Poiché le barriere si incontrano per  $u_K = l_K$ , risolvendo un'equazione di secondo grado si ottiene:

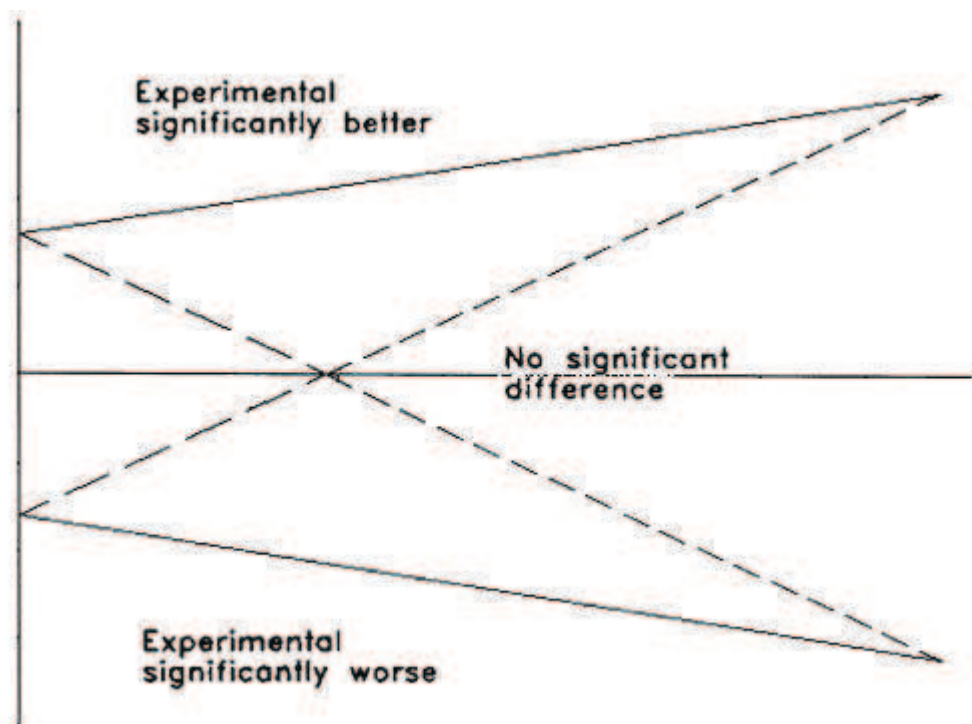
$$V_{max}^c = \left[ \sqrt{\frac{4 \times 0.583^2}{K} + 16ac} - \frac{2 \times 0.583}{\sqrt{K}} \right]^2 \frac{1}{16\delta^2} \quad (3.3)$$

dove  $V_{max}^c$  indica il nuovo valore di  $V_{max}$  calcolato utilizzando la correzione di continuità. Il valore di  $V_{max}^c$  risulterà inferiore rispetto a quello calcolato senza la correzione di continuità  $V_{max} = a/c$ .

### 3.3 Test triangolare inverso e triangolare doppio

Per avere maggiore e predeterminata potenza nell'indagare sull'inferiorità di un trattamento si utilizza il test triangolare inverso. Un test di questo tipo verifica l'ipotesi nulla  $H_0 : \theta = 0$  contro l'alternativa  $H_1 : \theta < 0$ . I vincoli imposti per soddisfare i requisiti di potenza in un test di questo tipo sono  $C^-(-\delta) = 1 - \beta$  e  $C^-(0) = \alpha/2$ . Questo test è utilizzato quando l'ipotesi alternativa di superiorità non è di interesse. Le barriere sono esattamente l'opposto di quelle del test triangolare singolo con ipotesi alternativa  $\theta > 0$ , cioè  $u'_k = -l_k$  e  $l'_k = -u_k$ . La regola di arresto viene applicata in modo speculare a quella del test singolo visto in precedenza.

Un test triangolare doppio consiste nella combinazione di un test triangolare con uno triangolare inverso, la zona di continuazione è definita da 4 barriere,  $u_k, l_k, u'_k, l'_k$ . È utilizzato quando si vuole verificare un'ipotesi con alternativa bilaterale del tipo  $H_1 : \theta = \pm\delta$ . Diversamente da quello singolo,



**Figura 3.2:** Test triangolare doppio disegnato sul piano  $(V, S)$ . Fonte: Whitehead (1997, Figura 4.26).

il test triangolare doppio richiede  $C(0) = \alpha$  e  $C^+(\delta) = C^-(\delta) = 1 - \beta$ . Esso infatti desidera avere una predefinita potenza nell'individuare sia inferiorità che superiorità del trattamento sperimentale.

Il test triangolare doppio continua finché non si raggiunge una delle tre regioni di arresto comune ad entrambi i test. Le tre regioni di arresto porgono evidenza in favore di superiorità del trattamento sperimentale se  $S_k \geq u_k$ , inferiorità del trattamento sperimentale se  $S_k \leq l'_k$  e uguaglianza dei due trattamenti se  $l_k \geq S_k \geq u'_k$ , come mostrato in Figura 3.2. L'ultima disuguaglianza può essere verificata solo dal punto in cui le barriere  $u'_k$  e  $d_k$  si incrociano, in poi. In assenza di uno dei tre casi il test continua finché le soglie non si incontrano, cioè fino al raggiungimento di  $V_{max}$ .

### 3.4 Analisi dell'esperimento

Una volta concluso l'esperimento, è possibile confrontare i risultati ottenuti con quelli di altri test, e valutare la credibilità del risultato ottenuto.

È interessante notare che in un test a campione fisso l'informazione necessaria per condurre l'esperimento è un valore fisso determinato dai requisiti di potenza e dal valore di  $\delta$ . In un test triangolare non si può sapere in via preventiva quale sia il livello di informazione necessario per concludere l'esperimento, dato che non si può prevedere in quale analisi si verifichi la regola d'arresto. Si conosce solamente  $V_{max}$  che rappresenta il livello di informazione necessario nel caso peggiore, ossia quando si prosegue fino all'ultima analisi. Si definisce quindi  $V^*$ , il valore dell'informazione al momento della conclusione l'esperimento, ed è una variabile casuale con massimo nel punto  $V_{max}$ . Ci si attende che in un test sequenziale, la quantità di informazione utilizzata per concludere l'esperimento sia inferiore a quella di un test a campione fisso. Essendo  $V^*$  una variabile casuale, è possibile derivarne la distribuzione, calcolarne il valore atteso, la mediana e i percentili.

Per avere un'idea di come derivare la distribuzione di  $V^*$  si definiscano

$$\begin{aligned} f_i &= P(S_k \in (l_k, u_k), k = 0, \dots, (i-1) \text{ e } S_i \leq l_i | \theta = 0) \\ g_i &= P(S_k \in (l_k, u_k), k = 0, \dots, (i-1) \text{ e } S_i \leq u_i | \theta = 0) \\ h_i &= P(S_k \in (l_k, u_k), k = 0, \dots, (i-1) | \theta = 0) \end{aligned}$$

con  $i = 1, \dots, K$ , ipotizzando  $S_0 = 0$ ,  $l_0 = -\infty$  e  $u_0 = \infty$ . La quantità  $f_i$  definisce la probabilità di arrestare l'esperimento dopo aver violato la soglia inferiore all' $i$ -esima analisi e non prima e il valore di  $(h_i - g_i)$  è la probabilità di arrestare l'esperimento dopo aver violato la soglia superiore all' $i$ -esima analisi e non prima. Il valore atteso dell'informazione  $V^*$  al termine dell'esperimento è dato da:

$$E(V^*) = (f_1 + h_1 - g_1)V_1 + \dots + (f_K + h_K - g_K)V_K$$

Dalle stesse probabilità si può ricavare anche la mediana,  $MED(V^*; \theta)$ , definita da

$$P(V^* \leq MED(V^*; \theta); \theta) = 0.5.$$

In modo analogo si possono calcolare i percentili. Per approfondimenti sul calcolo della distribuzione di  $V^*$ , si veda Whitehead (1997, Paragrafo 4.11.2).

Quando un esperimento sequenziale viene concluso, si ha evidenza in favore di una delle due ipotesi. È possibile avere ulteriore riprova dell'evidenza che l'esperimento sostiene, esprimendo i risultati in termini di livello di significatività osservata, stime puntuali e intervalli di confidenza. L'interpretazione di tali misure è la stessa adottata per i test a campione fisso, ma nell'ambito degli esperimenti sequenziali non esiste un unico approccio per calcolarli. Sono stati proposti vari metodi, che possono portare ad ottenere risultati differenti a partire dallo stesso insieme di dati.

Di seguito tali quantità verranno espone facendo riferimento alle definizioni di Whitehead (1997, Paragrafi 5.2-5.3). La funzione  $p$ -value,  $P(\theta)$  è definita come la probabilità, quando il valore del parametro è  $\theta$ , di ottenere evidenza in favore dell'ipotesi di superiorità del trattamento sperimentale forte quanto quella ottenuta con l'insieme di dati osservati o più forte. Per calcolare tale funzione è richiesto quindi un metodo che consenta di confrontare e ordinare tutte le possibili realizzazioni dell'esperimento. L'approccio convenzionale utilizza l'ordinamento di Fairbanks e Madsen (1982). In base a questo ordinamento, un esperimento che termina con le statistiche  $(s^{(1)}, v^{(1)})$  fornisce maggiore evidenza di un esperimento che termina con  $(s^{(2)}, v^{(2)})$  se  $s^{(1)} > s^{(2)}$  e  $v^{(1)} = v^{(2)}$ , o se  $s^{(1)}$  supera la barriera superiore, e  $v^{(1)} < v^{(2)}$  oppure se  $s^{(2)}$  supera la barriera inferiore, e  $v^{(1)} > v^{(2)}$ . Ad esempio, si supponga di condurre un esperimento con  $K = 2$  analisi con valori finali delle statistiche  $(s^*, v^*)$ . Se  $v^* = V_1$  allora:

$$P(\theta) = P(S_1 \geq s^*) = 1 - \Phi\{(b^* - \theta V_1)/\sqrt{V_1}\}$$

mentre se  $v^* = V_2$

$$\begin{aligned} P(\theta) &= P(S_1 \geq u_1 \text{ o } \{S_1 \in (l_1, u_1) \text{ e } S_2 \geq b^*\}) \\ &= \Phi(-u_1/\sqrt{V_1} + \theta\sqrt{V_1}) + \Phi_2(-l_1/\sqrt{V_1} + \theta\sqrt{V_1}, -(b^* - \theta V_2)/\sqrt{V_2}, \rho) \\ &\quad - \Phi_2(-u_1/\sqrt{V_1} + \theta\sqrt{V_1}, -(b^* - \theta V_2)/\sqrt{V_2}, \rho), \end{aligned}$$

Una volta definita la funzione  $p$ -value è possibile calcolare il livello di significatività per un test con alternativa unilaterale:

$$p_+ = P(0)$$

e per un test con alternativa bilaterale:

$$p = 2 \min \{P(0), 1 - P(0)\}.$$

La funzione  $p$ -value è utilizzata come informazione aggiuntiva sui risultati ottenuti con un esperimento sequenziale. La conclusione di superiorità del trattamento sperimentale è supportata se si ottiene  $P(0) \leq \alpha$ . Viceversa,  $P(0) \geq 1 - \alpha$  è a sostegno dell'ipotesi di inferiorità del trattamento sperimentale. Tramite la funzione  $p$ -value è inoltre possibile definire intervalli di confidenza per  $\theta$ . Ad esempio, si definiscano  $\theta_L$  e  $\theta_U$  tali che  $P(\theta_L) = 0.025$  e  $P(\theta_U) = 0.975$ . Allora,  $P(\theta \leq \theta_L(D)) = 0.025$  e  $P(\theta \leq \theta_U(D)) = 0.975$ . Ne segue che  $P(\theta_L(D) \leq \theta \leq \theta_U(D)) = 0.95$ , (cfr. Whitehead 1997, Paragrafo 5.2) dove  $\theta(D)$  è una variabile casuale che rappresenta il valore di  $\theta$  rispetto a un potenziale dataset,  $D$ . Perciò  $(\theta_L, \theta_U)$ , calcolati utilizzando i dati osservati formano un intervallo di confidenza di livello 95% per  $\theta$ .

### **Esempio: Test triangolare con risposte binarie**

Quando le risposte dei pazienti possono assumere solo due valori, siamo nel caso di risposte binarie. Solitamente uno dei due valori è rappresentativo di un “successo” e l'altro di un “insuccesso”, come per esempio il manifestarsi o meno di effetti collaterali. Si supponga di analizzare risposte indipendenti

provenienti da pazienti trattati con un trattamento sperimentale e con uno di controllo. Le risposte provengono dalle variabili casuali  $X \sim \text{Bin}(mk, p_C)$  e  $Y \sim \text{Bin}(mk, p_E)$ , in riferimento al trattamento di controllo e sperimentale rispettivamente. Si supponga di svolgere la  $k$ -esima analisi, e quindi di avere a disposizione  $2mk$  risposte osservate.

In generale, negli esperimenti comparativi, si hanno osservazioni provenienti da due diverse densità, ovvero  $f(x; \psi_1, \eta)$  e  $f(x; \psi_2, \eta)$  dove  $\eta$  è un parametro di disturbo comune ignoto. Il parametro di interesse può essere ad esempio  $\theta = \frac{1}{2}(\psi_1 - \psi_2)$  e i parametri di disturbo sono ad esempio  $\phi = \frac{1}{2}(\psi_1 + \psi_2)$  ed  $\eta$ .

In questo esempio  $\eta$  non è presente, e il parametro di disturbo deriva dalla riparametrizzazione utilizzata per esplicitare  $\theta$  e viene indicato con  $\lambda$ . Il parametro di interesse che si desidera utilizzare per misurare la differenza tra i due trattamenti è il *log odds ratio*:

$$\theta = \log \left\{ \frac{p_S(1 - p_C)}{p_C(1 - p_S)} \right\}.$$

Si noti che il suo valore è nullo sotto l'ipotesi nulla di uguaglianza dei due trattamenti.

Si riparametrizza il modello utilizzando  $\psi_1 = 2 \log\{p_E(1 - p_E)\}$  e  $\psi_2 = 2 \log\{p_C(1 - p_C)\}$  in modo che

$$\theta = \log \left\{ \frac{p_E(1 - p_C)}{p_C(1 - p_E)} \right\} \quad \text{e} \quad \lambda = \log \left\{ \frac{p_E p_C}{(1 - p_C)(1 - p_E)} \right\}$$

Il parametro di disturbo è  $\lambda$ .

La log verosimiglianza per il modello utilizzato per il trattamento sperimentale utilizzando i dati fino all'analisi  $k$  è:

$$\begin{aligned} l(p_E) &= y_k \log p_E + (mk - y_k) \log(p_E), \quad \text{oppure} \\ l(\psi_1) &= \frac{1}{2} \psi_1 y_k - mk \log(1 + e^{1/2 \psi_1}). \end{aligned}$$

La log verosimiglianza congiunta, utilizzando la riparametrizzazione in  $\psi_1$  e

$\psi_2$  diventa

$$l(\psi_1, \psi_2) = \frac{1}{2}\psi_1 y_k - mk \log(1 + e^{\frac{1}{2}\psi_1}) + \frac{1}{2}\psi_2 x_k - mk \log(1 + e^{\frac{1}{2}\psi_2}).$$

Sapendo che  $\psi_1 = \lambda + \theta$  e  $\psi_2 = \lambda - \theta$  è possibile ottenere la log verosimiglianza in funzione dei parametri di interesse e di disturbo:

$$l(\theta, \lambda) = \frac{1}{2}y_k(\lambda + \theta) - mk \log(1 + e^{1/2(\lambda + \theta)}) + \frac{1}{2}x_k(\lambda - \theta) - mk \log(1 + e^{1/2(\lambda - \theta)})$$

Si può quindi procedere al calcolo di  $S_k = l_\theta(0, \hat{\lambda}_0^{(k)})$  fissando  $\lambda$  pari a  $\hat{\lambda}_0^{(k)}$ , e si ottiene:

$$S_k = \frac{y_k - x_k}{2}.$$

Per il calcolo di  $I_k$ , dato che si sta considerando il caso multiparametrico, si utilizza la formula:

$$\begin{aligned} I_k &= -\{l^{\theta\theta}(0, \hat{\lambda}_0^{(k)})\}^{-1} \\ &= l_{\theta\theta}(\theta, \hat{\lambda}_0^{(k)}) - \{j_{\theta\lambda}(\theta, \hat{\lambda}_0^{(k)})\}\{l_{\lambda\lambda}(\theta, \hat{\lambda}_0^{(k)})\}^{-1}\{l_{\theta\lambda}(\theta, \hat{\lambda}_0^{(k)})\}' \\ &= \frac{(x_k + y_k)(2mk - x_k - y_k)}{8mk}. \end{aligned}$$

Una volta ottenuti  $S_k$  e  $V_k$  si possono derivare le barriere

$$\begin{aligned} u_k &= \frac{2}{\delta} \log\left(\frac{1}{\alpha}\right) + \frac{1}{4}\delta I_k \\ l_k &= -\frac{2}{\delta} \log\left(\frac{1}{\alpha}\right) + \frac{3}{4}\delta I_k \end{aligned}$$

dove  $\delta$  è il valore del parametro di interesse sotto l'ipotesi alternativa. Le barriere assumono questa forma ipotizzando  $\frac{1}{2}\alpha = \beta$  e monitoraggio continuo.



# Capitolo 4

## Estensione del test di Whitehead

### 4.1 Introduzione

L'idea alla base di questo lavoro è applicare i metodi asintotici di verosimiglianza di ordine superiore al test di Whitehead per esperimenti comparativi. Tale applicazione è suggerita dal fatto che il test triangolare utilizza quantità basate sulla verosimiglianza e sfrutta l'approssimazione normale per la distribuzione della *score*. Whitehead (1997, Paragrafo 3.9) suggerisce la possibilità di utilizzare la versione log rapporto di verosimiglianza in luogo della statistica *score* (Whitehead, 1997, Paragrafo 3.9.1) e di ricorrere a trasformazioni normalizzate per migliorare l'approssimazione normale (Whitehead, 1997, Paragrafo 3.9.4).

Nel Paragrafo 2 di questo capitolo verranno introdotti i metodi asintotici di ordine superiore basandosi sui volumi di Pace e Salvani (1996) e Severini (2000). Nei Paragrafi 3 e 4 verrà spiegata l'intuizione alla base del lavoro svolto, i risultati attesi e l'organizzazione della ricerca per verificarli. I risultati ottenuti saranno esposti nel Paragrafo 5.

## 4.2 I metodi asintotici di ordine superiore

La teoria asintotica di ordine superiore si basa su risultati relativamente recenti, successivi al 1980. Come indicato dal nome, i metodi di ordine superiore permettono di ottenere approssimazioni asintotiche per le quantità di verosimiglianza con errore asintotico più piccolo rispetto alle approssimazioni del primo ordine. Pertanto, tali metodi consentono di raffinare alcuni dei risultati del primo ordine.

Si consideri un modello statistico parametrico regolare con generico parametro  $\psi \in \Psi \subseteq \mathbb{R}^p$ . Si indichi con  $s(x)$  la statistica sufficiente minimale di tale modello e si supponga che possa essere rappresentata in funzione di  $(\hat{\psi}, a)$ , dove  $a$  è una statistica ancillare, ossia con distribuzione indipendente da  $\psi$  e tale che la relazione che lega  $s(x)$  e  $(\hat{\psi}, a)$  sia biunivoca. In questo caso la funzione di log verosimiglianza,  $l(\psi)$ , può essere riscritta come  $l(\psi; \hat{\psi}, a)$ , per enfatizzare il fatto che è funzione dei dati tramite la statistica sufficiente  $(\hat{\psi}, a)$ , oltre che del parametro  $\psi$ .

I metodi asintotici di ordine superiore basano molti dei loro risultati sull'utilizzo di un'approssimazione della densità dello stimatore di massima verosimiglianza condizionata alla statistica ancillare  $a$ , altrimenti chiamata formula  $p^*$  di Barndorff-Nielsen (1980, 1983).

La formula  $p^*$  ha la seguente espressione:

$$p^*(\hat{\psi}; \psi|a) = c(\psi, a) |j(\hat{\psi}; \hat{\psi}, a)|^{1/2} \exp\{l(\psi; \hat{\psi}, a) - l(\hat{\psi}; \hat{\psi}, a)\} \quad (4.1)$$

con  $c(\psi, a)$  costante di normalizzazione. In generale,  $p^*$  approssima la densità condizionata esatta di  $\hat{\psi}|a$  con errore di ordine  $O(n^{-3/2})$ , cioè:

$$p_{\hat{\psi}|A=a}(\hat{\psi}; \psi, a) = p^*(\hat{\psi}; \psi|a) \{1 + O(n^{-3/2})\}.$$

Nel caso di famiglie di posizione e scala, con  $\hat{\psi} = (\hat{\mu}, \hat{\sigma})$  ed  $a = (a_1, \dots, a_n)$ ,  $p^*$  dà esattamente la distribuzione di  $\hat{\psi}|a$  ove  $a$  è la statistica di configurazione, con  $a_i = (y_i - \hat{\mu})/\hat{\sigma}$ ,  $i = 1, \dots, n$ . Altri casi di esattezza non verranno qui

trattati. Per approfondimenti si rimanda a Pace e Salvan (1996, Paragrafo 11.2) e Severini (2000, Paragrafi 6.3, 6.6.5).

### 4.2.1 Versioni modificate di $Z$ e $Z_p$

Tramite l'utilizzo della formula  $p^*$  è possibile modificare alcune quantità calcolate con i metodi asintotici del primo ordine allo scopo di ottenere risultati di approssimazione più accurati. Tali perfezionamenti riguardano in particolare le distribuzioni nulle di  $W$ ,  $W_e$ ,  $W_u$  e  $Z$ , definiti nel Paragrafo 2.2.1, anche nelle analoghe versioni per modelli con parametri di disturbo. In questo paragrafo ci si concentrerà solamente sulla versione modificata di  $Z$  e  $Z_p$ . Verrà utilizzato il concetto di ordine di una successione, indicato con il simbolo  $O(\cdot)$  nel caso deterministico e con  $O_p(\cdot)$  nel caso stocastico (cfr. Pace e Salvan, 1997, Paragrafo 3.4.1).

Si consideri innanzitutto la seguente trasformazione per una variabile casuale con distribuzione approssimata normale,  $X$ , con errore  $O_p(1)$  e densità della forma

$$\bar{c}h(x/\sqrt{n})\phi(x) \quad (4.2)$$

dove  $h$  è di ordine  $O(1)$ ,  $h(0) = 1$  e  $\bar{c}$  è una costante del tipo  $\bar{c} = 1 + O(n^{-1})$ . È possibile dimostrare (cfr. Severini, 2000, Paragrafo 7.2) che una trasformazione di  $X$  è data da:

$$X^* = X - \frac{\log h(X/\sqrt{n})}{X}.$$

La variabile casuale  $X^*$  ha distribuzione normale standard con errore di ordine  $O_p(n^{-3/2})$ .

Si consideri quindi il test log rapporto di verosimiglianza,  $Z$ , nel caso monoparametrico:

$$Z = \text{sgn}(\hat{\psi} - \psi) \{2(l(\hat{\psi}) - l(\psi))\}^{1/2}.$$

L'approssimazione  $p^*$  è data da

$$c j(\hat{\psi})^{1/2} \exp\{l(\psi) - l(\hat{\psi})\}$$

dove  $l(\psi) \equiv l(\psi; \hat{\psi}, a)$  è visto come funzione di  $\hat{\psi}$  per  $a$  fissata, e  $c$  può essere scritto come  $c = \bar{c}/\sqrt{2\pi}$  con  $\bar{c} = 1 + O(n^{-1})$ . Si noti che

$$\frac{1}{2}Z^2 = l(\hat{\psi}) - l(\psi) \quad (4.3)$$

perciò la densità di  $Z$  può essere approssimata da

$$c j(\hat{\psi})^{1/2} \exp\left\{-\frac{1}{2}z^2\right\} \left| \frac{\partial \hat{\psi}}{\partial z} \right|$$

dove  $j(\hat{\psi})$  è visto come una funzione di  $z$ . Utilizzando la (4.3), ed indicando con  $l_{;\hat{\psi}}(\cdot)$  la derivata prima della log verosimiglianza, rispetto alla stima di massima verosimiglianza di  $\psi$ , segue che:

$$z \frac{\partial z}{\partial \hat{\psi}} = l_{;\hat{\psi}}(\hat{\psi}) - l_{;\hat{\psi}}(\psi);$$

perciò

$$\left| \frac{\partial \hat{\psi}}{\partial z} \right| = \frac{|z|}{|l_{;\hat{\psi}}(\hat{\psi}) - l_{;\hat{\psi}}(\psi)|}.$$

Un'ulteriore approssimazione per la densità di  $Z$ , considerando  $\hat{\psi}$  come funzione di  $z$  è quindi

$$\bar{c} \frac{j(\hat{\psi})^{1/2} |z|}{|l_{;\hat{\psi}}(\hat{\psi}) - l_{;\hat{\psi}}(\psi)|} \phi(z)$$

con  $\bar{c} = 1 + O(n^{-1})$ . Si nota che questa approssimazione è della forma (4.2) con

$$h(z/\sqrt{n}) = \frac{\bar{c} j(\hat{\psi})^{1/2} |z|}{|l_{;\hat{\psi}}(\hat{\psi}) - l_{;\hat{\psi}}(\psi)|}$$

e  $h(0) = 1$ , essendo che  $z = 0$  corrisponde a  $\hat{\psi} = \psi$ . La versione modificata

di  $Z$  pertanto è data da

$$Z^* = Z + \frac{1}{Z} \log(U/Z),$$

con

$$U = j(\hat{\psi})^{-1/2} \{l_{;\hat{\psi}}(\hat{\psi}) - l_{;\hat{\psi}}(\psi)\}.$$

Il test  $Z^*$  ha distribuzione  $N(0, 1)$  con errore di ordine  $O_p(n^{-3/2})$ . Il segno di  $U$  è lo stesso di  $Z$  perciò  $U/Z$  non può essere negativo.

Nel caso in cui il parametro complessivo sia suddiviso in due blocchi di componenti,  $(\theta, \lambda)$ , indicative rispettivamente del parametro scalare di interesse e di quello di disturbo di dimensione  $p - 1$ , è possibile calcolare la versione modificata di  $Z_p$ , avente distribuzione nulla  $N(0, 1)$  con ordine di errore  $O_p(n^{-3/2})$ . I passaggi svolti per ottenere tale quantità si possono trovare in Severini (2000, Paragrafo 7.4). Assumendo che la statistica sufficiente minimale sia esprimibile nella forma  $(\hat{\theta}, \hat{\lambda}, a)$  con  $a$  ancillare, la versione modificata di  $Z_p$  è:

$$Z_p^* = Z_p + \frac{1}{Z_p} \log(U_p/Z_p) \quad (4.4)$$

con

$$U_p = \frac{\begin{vmatrix} l_{;\hat{\psi}}(\hat{\psi}) - l_{;\hat{\psi}}(\hat{\psi}_\theta) \\ l_{\lambda;\hat{\psi}}(\hat{\psi}_\theta) \end{vmatrix}}{(|j_{\lambda\lambda}(\hat{\psi}_\theta)|^{1/2} |j(\hat{\psi})|^{1/2})}, \quad (4.5)$$

dove  $j(\hat{\psi})$  è la matrice di informazione osservata per  $\psi$  calcolata in  $\hat{\psi} = (\hat{\theta}, \hat{\lambda})$ ,  $j_{\lambda\lambda}(\hat{\psi}_\theta)$  è il blocco  $(\lambda, \lambda)$  della matrice di informazione osservata calcolata in  $\hat{\psi}_\theta = (\theta, \hat{\lambda}_\theta)$  e  $l_{\lambda;\hat{\psi}}(\hat{\psi}_\theta) = \partial l_{;\hat{\psi}}(\hat{\psi}_\theta) / \partial \lambda$ .

La teoria suggerisce pertanto che sostituire una quantità calcolata con metodi asintotici del primo ordine, con il suo equivalente calcolato con metodi asintotici di ordine superiore, consente di ridurre l'errore di approssimazione. Ad esempio, sostituendo la statistica  $Z_p$  con il suo equivalente  $Z_p^*$ , l'errore di approssimazione utilizzando la distribuzione normale passa da  $O_p(n^{-1/2})$  a  $O_p(n^{-3/2})$ .

### 4.3 L'idea di base

Essendo il test di Whitehead basato sull'approssimazione normale per la distribuzione della *score*, è possibile che i risultati ottenuti non siano molto accurati soprattutto nel caso di piccoli campioni. A tal proposito si noti che in base alla teoria, la convergenza di  $Z_p^*$  alla propria distribuzione asintotica è più veloce di quanto accade per i test che utilizzano la teoria asintotica del primo ordine. Si è pensato quindi di modificare il test di Whitehead, in modo da renderlo più preciso nei casi in cui si hanno a disposizione poche osservazioni.

Il ragionamento parte dal fatto che il test di Whitehead utilizza la statistica  $S_k(\theta) = Z_{up,k}\sqrt{V_k(\theta)}$  calcolata in  $\theta = 0$  e che le statistiche  $Z_p$  e  $Z_{up}$  sono equivalenti a livello asintotico. Si ha quindi che una sostituzione della statistica  $S_k(\theta)$  con  $Z_{p,k}\sqrt{V_k(\theta)}$  non provoca variazioni a livello asintotico, dato che entrambe le quantità hanno distribuzione asintotica nulla  $N(0, V_k(0))$ . Tale equivalenza è sostenuta anche da Whitehead (1997, p. 60). È possibile inoltre sostituire ulteriormente  $Z_{p,k}\sqrt{V_k(\theta)}$  con la sua forma equivalente  $Z_{p,k}^*\sqrt{V_k(\theta)}$ , calcolata in  $\theta = 0$ , che si ritiene possa fornire una migliore approssimazione asintotica e ha distribuzione asintotica  $N(0, V_k(0))$ . Dall'ultima sostituzione ci si attende una maggior accuratezza dei risultati, particolarmente in caso di bassa numerosità campionaria. Se questo fosse verificato, per piccoli campioni, il test di Whitehead modificato sarebbe migliore di quello classico.

Si desidera quindi verificare se la sostituzione di  $S_k(\theta)$  con  $Z_{p,k}^*\sqrt{V_k(\theta)}$  produce effettivamente i miglioramenti attesi.

### 4.4 Struttura della ricerca

Per provare la validità dell'intuizione del paragrafo precedente, si sono utilizzate simulazioni Monte Carlo con 50000 replicazioni. Avvalendosi del Software R (R Development Core Team, 2012) e prendendo come riferimento di confronto Whitehead (1997), è stato applicato il test triangolare in tre differenti versioni. La prima utilizza come statistica test quella scelta da

Whitehead, cioè  $S_k(\theta)$ , la seconda utilizza  $Z_{p,k}\sqrt{V_k(\theta)}$  e la terza utilizza  $Z_{p,k}^*\sqrt{V_k(\theta)}$ , tutte calcolate in  $\theta = 0$ .

Per confrontare tra loro i tre tipi test, sono stati stimati, per ognuno, la probabilità di errore di I tipo dell'esperimento nel suo complesso e il valore atteso dell'informazione osservata necessaria per arrivare ad una conclusione,  $E(V^*)$ . Da quest'ultimo, tramite l'espressione di  $V_{max}$ , si può derivare il valore atteso della numerosità campionaria necessaria per concludere l'esperimento,  $E(mk^*)$ . Questo valore viene poi espresso come percentuale rispetto alla numerosità campionaria necessaria di un test a numerosità campionaria fissata per verificare le stesse ipotesi con le stesse probabilità di errore di I e II tipo, e verrà denotato con il simbolo  $E(mk^*)\%$ . La probabilità di errore di I tipo,  $\hat{\alpha}$  viene utilizzata come indice della bontà dell'approssimazione. Più il valore di  $\hat{\alpha}$  è vicino al suo valore nominale,  $\alpha$ , più il metodo è preciso. Il valore di  $E(mk^*)\%$  è invece indice dell'efficacia di un metodo sequenziale in termini di riduzione della numerosità campionaria rispetto ad un test a numerosità campionaria fissata. A parità di precisione, viene considerato preferibile il test che minimizza  $E(mk^*)\%$  stimato.

Le assunzioni fatte per replicare il test triangolare sono le seguenti. Si hanno a disposizione due gruppi di osservazioni indipendenti provenienti da pazienti assegnati a due differenti trattamenti:  $x_1, \dots, x_{mK}$  e  $y_1, \dots, y_{mK}$  con generica distribuzione  $f(x; \theta_C, \lambda)$  e  $f(y; \theta_E, \lambda)$ . Il pedice  $C$ , viene utilizzato per contraddistinguere il modello utilizzato per il trattamento di controllo, mentre il pedice  $E$  viene utilizzato per il trattamento sperimentale. In entrambi i modelli il vettore di parametri è suddiviso in una componente di interesse scalare,  $\theta_C$  o  $\theta_E$ , e una di disturbo comune ad entrambi,  $\lambda$ . Si desidera verificare l'uguaglianza tra i due trattamenti, ovvero tra i due modelli. Si definisce quindi il parametro scalare  $\theta$  che è una riparametrizzazione dei due parametri  $\theta_C$  e  $\theta_E$ , tale che valga 0 quando  $\theta_E = \theta_C$ .

L'ipotesi nulla di uguaglianza dei due trattamenti viene verificata sia nel

caso di alternativa bilaterale che nel caso di alternativa unilaterale, ossia:

$$\begin{aligned} H_0 : \theta = 0 & \quad \text{contro} \quad H_1 : \theta \neq 0 \quad \text{e} \\ H_0 : \theta = 0 & \quad \text{contro} \quad H_1 : \theta > 0. \end{aligned}$$

Nel primo caso viene utilizzato un test triangolare singolo, mentre nel secondo caso un test triangolare doppio. A tal proposito si ricorda che per un test triangolare singolo si impone  $C(0) = C^+(0) = \alpha/2$  e per un test triangolare doppio  $C(0) = C^+(0) + C^-(0) = \alpha$ , dove  $C(0)$  è la probabilità di rifiutare l'ipotesi nulla quando questa è vera. Il livello di significatività scelto è  $\alpha = 0.05$  e la potenza  $\beta = \alpha/2 = 0.025$ .

Sono stati considerati esperimenti con  $K$  analisi, ognuna svolta al raggiungimento di  $m$  osservazioni aggiuntive per ogni campione. Le analisi sono quindi equispaziate. La numerosità campionaria per la  $k$ -esima analisi è  $2mk$ , e la numerosità campionaria totale dell'esperimento è  $2mK$ . Si è scelto di confrontare i tre test per diverse numerosità campionarie totali, variando il valore di  $K$  e di  $m$ . Per  $K$  sono stati scelti i valori 2, 4, 6, 8 e per  $m$  i valori 5, 10, 15, 20, 50.

In fase di progettazione dell'esperimento la prima cosa da fare è definire le soglie, ossia calcolare i valori di  $a$  e  $c$ , e definire il valore dell'informazione attesa massima,  $V_{max}$ . Normalmente in base a  $V_{max}$  viene definita anche la numerosità campionaria massima, e quindi  $m$  e  $K$ . Nel nostro caso però, si vuole imporre una bassa numerosità campionaria, perciò i valori di  $m$  e  $K$  sono definiti a priori. Ne segue che il valore di  $\delta$  verrà imposto dalla scelta della numerosità campionaria totale, ossia: viene calcolato il valore di  $V_{max}$  sotto  $H_0$  utilizzando come numero di osservazioni il numero massimo di osservazioni disponibili per un campione,  $mK$ ; si impone che  $V_{max}$  corrisponda al vertice del triangolo, cioè al punto d'incontro delle barriere:

$$V_{max} = \frac{8}{\delta^2} \log\left(\frac{1}{\alpha}\right) \quad (4.6)$$

e l'unica quantità ignota è  $\delta$  che viene quindi definita tramite quest'equazione.



È possibile ora calcolare le costanti  $a$  e  $c$  utilizzando le formule (3.1).

Il passo successivo è quello di considerare la correzione di continuità delle soglie, che ha l'effetto di ridurre l'informazione massima necessaria per l'esperimento. Viene quindi calcolata tramite la (3.3) l'informazione massima che verrà effettivamente utilizzata. Si noti che la correzione del valore dell'informazione massima, comporta anche una correzione nella numerosità campionaria totale  $2mK$ . Si è deciso di seguire l'approccio indicato da Jennison (2000, Paragrafo 4.5), ossia di mantenere inalterato il valore di  $K$  e di modificare il valore di  $m$ . Per via della correzione di continuità, quindi, il valore di  $m$  effettivamente utilizzato sarà inferiore a quello definito inizialmente. Tale riduzione non influisce in termini di confrontabilità tra i metodi dato che è uguale per tutti e tre i test.

## 4.5 Modelli e simulazioni

Sono stati considerati due modelli, uno per dati continui e uno per dati discreti. Per i dati continui è stata utilizzata la distribuzione normale con varianza ignota, mentre per i dati discreti è stata utilizzata la distribuzione binomiale, in due alternative contraddistinte da diversi parametri di interesse.

### 4.5.1 Il modello normale con varianza ignota

In questo esempio, vengono considerati dati normali generati dalle variabili casuali  $X_k \sim N(\mu_C, \sigma^2)$  e  $Y_k \sim N(\mu_E, \sigma^2)$  con parametri ignoti. Le osservazioni provenienti dalle due variabili casuali hanno uguale numerosità. Come misura della differenza tra i due trattamenti si è inizialmente considerato il parametro di interesse  $\theta = \mu_E - \mu_C$  e il parametro di disturbo è stato posto  $\lambda = (\mu_C, \sigma^2)'$ . Per non appesantire la notazione, dato che i parametri di disturbo non sono stati sottoposti a nessuna riparametrizzazione, vengono denotati di seguito con il loro simbolo originale e non con il simbolo  $\lambda$ .

La funzione di log verosimiglianza congiunta, per  $k = 1, \dots, K$ , è

$$l(\theta, \mu_C, \sigma^2; x_k, y_k) = -mk \log \sigma^2 - \frac{mk}{2\sigma^2} \{ (s_{y_k}^2 + s_{x_k}^2) + (\bar{y}_k - \theta - \mu_C)^2 + (\bar{x}_k - \mu_C)^2 \}$$

dove  $s_{y_k}^2 = \sum_{i=1}^{mk} (y_{k,i}^2 - \bar{y}_k^2)/mk$ ,  $x_k$  e  $y_k$  sono le osservazioni campionarie riferite alla  $k$ -esima analisi e  $\bar{x}_k$  e  $\bar{y}_k$  sono le medie campionarie calcolate alla  $k$ -esima analisi. In questo caso il parametro generale  $\psi = (\theta, \lambda)$  ha dimensione  $p = 3$ . Le stime di massima verosimiglianza dei parametri sono  $\hat{\theta}^{(k)} = \bar{y}_k - \bar{x}_k$ ,  $\hat{\mu}_C^{(k)} = \bar{x}_k$ ,  $\hat{\sigma}^{2(k)} = (s_{y_k}^2 + s_{x_k}^2)/2$ , e le stime di massima verosimiglianza vincolate, sotto  $H_0 : \theta = 0$ , sono  $\hat{\mu}_{C,0}^{(k)} = (\bar{y}_k + \bar{x}_k)/2$  e  $\hat{\sigma}_0^{2(k)} = (s_{y_k}^2 + s_{x_k}^2)/2 + ((\bar{y}_k - \bar{x}_k)/2)^2$ . Per questo modello l'informazione parziale attesa calcolata in corrispondenza del numero massimo di osservazioni è  $V_{max} = mK/(2\hat{\sigma}_0^{2(k)})$ . Tale valore però dipende dall'insieme totale dei dati e quindi non è possibile calcolarlo in fase di progettazione dell'esperimento. Questo impedisce di procedere oltre dato che  $V_{max}$  rappresenta il punto di partenza per calcolare le barriere e quindi per definire le altre quantità necessarie all'esperimento. Il parametro di interesse  $\theta = \mu_E - \mu_C$  in effetti viene sconsigliato da Whitehead (1997, Paragrafo 3.9.4), anche nel caso di varianza nota. Inoltre, tramite simulazioni, Facey (1992) dimostra che le proprietà di test sequenziali per dati normali basati su  $\theta = \mu_E - \mu_C$  non sono vicini alle previsioni teoriche. Si è quindi deciso di utilizzare come parametro di interesse quello consigliato da Whitehead (1997), cioè  $\theta = (\mu_E - \mu_C)/\sigma$ . I parametri di disturbo sono uguali a quelli del caso precedente:  $\mu_C$  e  $\sigma^2$ . La funzione di log verosimiglianza congiunta, per  $k = 1, \dots, K$ , è

$$l(\theta, \mu_C, \sigma^2; x_k, y_k) = -mk \log \sigma^2 - \frac{mk}{2\sigma^2} \{ (s_{y_k}^2 + s_{x_k}^2) + (\bar{y}_k - \theta\sigma - \mu_C)^2 + (\bar{x}_k - \mu_C)^2 \}.$$

Le stime di massima verosimiglianza dei parametri sono  $\hat{\theta}^{(k)} = (\bar{y}_k - \bar{x}_k)/\hat{\sigma}^{(k)}$ ,  $\hat{\mu}_C^{(k)} = \bar{x}_k$ ,  $\hat{\sigma}^{2(k)} = (s_{y_k}^2 + s_{x_k}^2)/2$ , e le stime di massima verosimiglianza vincolate a  $\theta = 0$  sono  $\hat{\mu}_{C,0}^{(k)} = (\bar{y}_k + \bar{x}_k)/2$  e  $\hat{\sigma}_0^{2(k)} = (s_{y_k}^2 + s_{x_k}^2)/2 + ((\bar{y}_k - \bar{x}_k)/2)^2$ . L'informazione di parziale attesa è  $V_k = mk/2$  e, calcolata per il numero massimo di osservazioni diventa  $V_{max} = mK/2$ . Risolvendo poi la

(4.6) è possibile trovare il valore di  $\delta$  che andrà sostituito nelle (3.1) per determinare le costanti  $a$  e  $c$ . Una volta trovati i valori di  $a$  e  $c$  è possibile tramite la (3.3) trovare il valore di  $V_{max}^c$ . Mantenendo inalterato il valore di  $K$ , e imponendo  $V_{max}^c = mK/2$ , è possibile trovare quale sarà il valore di  $m$  effettivamente utilizzato. Una volta definite le barriere, si può condurre l'esperimento. Vengono generati  $mK$  dati per entrambi i trattamenti da una distribuzione normale,  $X \sim N(\mu_C, \sigma^2)$  e  $Y \sim N(\mu_E, \sigma^2)$  con  $\sigma^2 = 4$ . Al fine di calcolare  $\hat{\alpha}$  entrambi i campioni vengono generati sotto l'ipotesi nulla di uguaglianza, perciò  $\mu_E = \mu_C = 0$ .

I valori delle quantità utilizzate sono

$$S_k = \frac{mk(\bar{y}_k - \bar{x}_k)}{2\hat{\sigma}_0^{(k)}}$$

$$I_k = mk \left( \frac{1}{2} - \frac{(\bar{y}_k - \bar{x}_k)^2}{16\hat{\sigma}_0^{2(k)}} \right)$$

$$Z_{p,k} = \text{sgn}(\hat{\theta}^{(k)}) \sqrt{2mk \log(\hat{\sigma}_0^{2(k)}/\hat{\sigma}^2(k))}.$$

Per calcolare  $Z_{p,k}^*$  viene utilizzata la log verosimiglianza riscritta in modo da essere funzione dei dati tramite le stime di massima verosimiglianza. Per ogni  $k = 1, \dots, K$ , ha la forma

$$l(\theta, \mu_C, \sigma^2; \hat{\theta}^{(k)}, \hat{\mu}_C^{(k)}, \hat{\sigma}^2(k)) = -mk \log \sigma^2 + \frac{mk}{\sigma^2} \left( -\hat{\sigma}^2(k) - \frac{\hat{\sigma}^2(k)\hat{\theta}^2(k)}{2} - \hat{\mu}_C^{2(k)} \right. \\ \left. - \hat{\mu}_C^{(k)}\hat{\theta}^{(k)}\hat{\sigma}^{(k)} + \hat{\theta}^{(k)}\hat{\sigma}^{(k)}\theta\sigma + \hat{\theta}^{(k)}\hat{\sigma}^{(k)}\mu_C + \hat{\mu}_C^{(k)}\theta\sigma \right. \\ \left. + 2\hat{\mu}_C^{(k)}\mu_C - \frac{\theta^2\sigma^2}{2} - \mu_C^2 - \mu_C\theta\sigma \right).$$

Utilizzando la (4.5) si ottiene la matrice al numeratore di  $U_p$

$$\begin{pmatrix} \frac{mk\hat{\sigma}^2(k)\hat{\theta}^{(k)}}{2\hat{\sigma}_0^{2(k)}} & 0 & -\frac{mk}{\hat{\sigma}^2(k)} + \frac{mk}{\hat{\sigma}_0^{2(k)}} \\ \frac{mk\hat{\sigma}^{(k)}}{\hat{\sigma}_0^{2(k)}} & \frac{2mk}{\hat{\sigma}_0^{2(k)}} & \frac{mk\hat{\theta}^{(k)}}{\hat{\sigma}_0^{2(k)}2\sqrt{\hat{\sigma}^2(k)}} \\ 0 & 0 & \frac{mk}{\hat{\sigma}_0^4(k)} \end{pmatrix},$$

la matrice  $j_{\lambda\lambda,k}(\hat{\psi}_0)$

$$\begin{pmatrix} \frac{2mk}{\hat{\sigma}_0^2(k)} & 0 \\ 0 & \frac{mk}{\hat{\sigma}_0^4(k)} \end{pmatrix},$$

e la matrice  $j_k(\hat{\psi})$

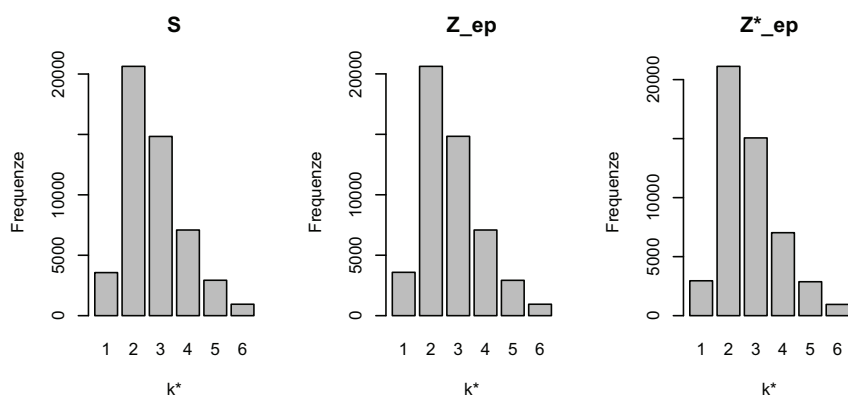
$$\begin{pmatrix} mk & \frac{mk}{\hat{\sigma}^{(k)}} & \frac{mk\hat{\theta}^{(k)}}{2\hat{\sigma}^2(k)} \\ \frac{mk}{\hat{\sigma}^{(k)}} & \frac{2mk}{\hat{\sigma}^2(k)} & \frac{mk\hat{\theta}^{(k)}}{2(\hat{\sigma}^2(k))^{3/2}} \\ \frac{mk\hat{\theta}^{(k)}}{2\hat{\sigma}^2(k)} & \frac{mk\hat{\theta}^{(k)}}{2(\hat{\sigma}^2(k))^{3/2}} & \frac{mk}{\hat{\sigma}^4(k)} \left(1 + \frac{\hat{\theta}^2(k)}{4}\right) \end{pmatrix}.$$

Si ottiene quindi

$$U_{p,k} = \frac{mk\hat{\sigma}^2(k)\hat{\theta}(k)}{\hat{\sigma}_0^2(k)\sqrt{2mk}}$$

e tramite la (4.4) è possibile calcolare  $Z_{p,k}^*$ .

Si procede poi a verificare se la regola di arresto è verificata. Nel caso in cui la statistica  $S_k$  non sia all'interno della zona di continuazione, l'esperimento termina, altrimenti si ripete il procedimento utilizzando ulteriori  $2m$  osservazioni. Se si arriva fino alla  $K$ -esima analisi, l'esperimento termina, e per costruzione ci sarà evidenza empirica in favore dell'ipotesi nulla o dell'alternativa. Una volta terminato l'esperimento, viene annotato tramite un contatore se l'ipotesi nulla è stata rifiutata. In questo modo, dividendo il valore finale del contatore per le 50000 replicazioni, è possibile ottenere una stima di  $\hat{\alpha}$  per l'esperimento nel suo complesso. Memorizzando inoltre in un vettore il numero  $k$  al quale termina ogni esperimento, è possibile ottenere la distribuzione di  $k^*$ , si veda ad esempio la Figura 4.1. Facendo la media di tutti i valori di  $k^*$  e moltiplicandolo per il valore di  $m$  utilizzato si ottiene la stima di  $E(mk^*)$ . Nel caso di test bilaterale, il numero di volte in cui viene rifiutata l'ipotesi nulla per superiorità o inferiorità del trattamento sperimentale vengono memorizzati in due contatori diversi e il numero totale di rifiuti è dato dalla loro somma. In questo modo è possibile verificare se l'errore di



**Figura 4.1:** Distribuzione delle frequenze di  $k^*$  per le statistiche  $S_k$ ,  $Z_{p,k}\sqrt{V_k}$  e  $Z_{p,k}^*\sqrt{V_k}$ , ottenuta con un test triangolare singolo con  $K=6$ ,  $m=10$ , 50000 replicazioni e osservazioni generate da  $N(0,4)$ .

primo tipo è simmetrico, cioè  $C^+(0) = C^-(0) = \alpha/2$ .

Lo stesso procedimento viene utilizzato per gli altri due test.

I risultati ottenuti dalle simulazioni sono riportati nelle Tabelle 4.1 e 4.2. L'errore di simulazione sul calcolo della probabilità di errore di primo tipo è stimato pari a 0.00097. Per calcolare  $E(mk^*)_{\%}$  è stata utilizzata la numerosità campionaria necessaria per un test a numerosità campionaria fissata unilaterale ossia

$$n_{f,1} = \frac{(\Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta))^2 2\sigma^2}{\delta^2}$$

e nel caso di test bilaterale si utilizza  $n_{f,2}$ , che si ottiene sostituendo  $\alpha/2$  al posto di  $\alpha$  nella formula di  $n_{f,1}$ .

Dalle tabelle si nota che in generale, per i tre test, al crescere della numerosità campionaria, i valori di  $\hat{\alpha}$  convergono al loro valore nominale (che è 0.025 nel caso di test unilaterale e 0.05 nel caso di test bilaterale). A parità di numero di osservazioni disponibili,  $mK$ , ad esempio con  $K = 2$  e  $m = 20$  o con  $K = 4$  e  $m = 10$ , il valore stimato di  $E(mk^*)_{\%}$  è decrescente al crescere del numero di analisi  $K$ . Inoltre il valore di  $E(mk^*)_{\%}$  tende a decrescere al crescere di  $mK$ .

Confrontando i tre test tramite i dati riportati nelle tabelle si nota che sia nel caso bilaterale che in quello unilaterale, per qualsiasi valore di  $K$  e  $m$ , il test  $Z_{p,k}^* \sqrt{V_k}$  restituisce valori di  $\hat{\alpha}$  più vicini al suo valore nominale. I valori di  $\hat{\alpha}$  riferiti ai test  $S_k$  e  $Z_{p,k} \sqrt{V_k}$  sono molto simili tra loro. La numerosità campionaria necessaria per concludere l'esperimento è sostanzialmente la stessa per tutti e tre i test. È possibile concludere quindi che il test che fornisce un'approssimazione migliore è  $Z_{p,k}^* \sqrt{V_k}$ .

La Tabella 4.3 riporta le percentuali di errore di primo tipo per superiorità e inferiorità del trattamento sperimentale conteggiate separatamente,

**Tabella 4.1:** Valori di  $\hat{\alpha}$  e numerosità campionaria attesa di un test triangolare singolo nelle tre varianti per osservazioni normali con varianza ignota. Osservazioni generate da  $N(0, 4)$ , parametro di interesse  $(\mu_E - \mu_C)/\sigma$ , 50000 simulazioni.

		Errore di I tipo			$E(mk^*)\%$		
		$S_k$	$Z_{p,k} \sqrt{V_k}$	$Z_{p,k}^* \sqrt{V_k}$	$S_k$	$Z_{p,k} \sqrt{V_k}$	$Z_{p,k}^* \sqrt{V_k}$
$K = 2$	$m = 5$	0.043	0.045	0.032	76.9	76.4	76.4
	10	0.034	0.034	0.028	81.6	81.6	81.2
	15	0.031	0.032	0.028	78.8	78.8	78.6
	20	0.030	0.030	0.027	80.6	80.6	80.4
	50	0.029	0.029	0.028	79.9	79.9	79.8
$K = 4$	$m = 5$	0.032	0.034	0.028	68.6	68.4	70.0
	10	0.029	0.029	0.026	69.7	69.7	70.2
	15	0.027	0.027	0.025	70.2	70.2	70.5
	20	0.027	0.027	0.025	70.1	70.1	70.2
	50	0.026	0.026	0.025	70.6	70.6	70.6
$K = 6$	$m = 5$	0.030	0.031	0.027	68.7	68.6	69.3
	10	0.026	0.026	0.025	67.9	67.9	68.2
	15	0.028	0.028	0.026	67.6	67.6	67.8
	20	0.026	0.026	0.025	67.5	67.5	67.7
	50	0.027	0.027	0.026	67.6	67.6	67.7
$K = 8$	$m = 5$	0.028	0.029	0.025	65.8	65.8	66.0
	10	0.026	0.026	0.025	65.8	65.8	65.9
	15	0.026	0.026	0.025	65.5	65.5	65.6
	20	0.026	0.026	0.025	65.6	65.6	65.7
	50	0.026	0.026	0.025	65.8	65.8	65.8

in riferimento al test triangolare doppio. Si può osservare che i risultati sono simmetrici, ossia la percentuale di rifiuto dell'ipotesi nulla per superiorità è molto simile a quella di rifiuto per inferiorità.

### 4.5.2 Il modello binomiale con parametro di interesse il *log odds ratio*

Si considerino dati binomiali generati dalle variabili casuali  $X_k \sim \text{Bin}(mk, p_C)$  e  $Y_k \sim \text{Bin}(mk, p_E)$ , in riferimento al trattamento di controllo e sperimenta-

**Tabella 4.2:** Valori di  $\hat{\alpha}$  e numerosità campionaria attesa di un test triangolare doppio nelle tre varianti per osservazioni nomali con varianza ignota. Osservazioni generate da  $N(0, 4)$ , parametro di interesse  $(\mu_E - \mu_C)/\sigma$ , 50000 simulazioni.

		Errore di I tipo			$E(mk^*)\%$		
		$S_k$	$Z_{p,k}\sqrt{V_k}$	$Z_{p,k}^*\sqrt{V_k}$	$S_k$	$Z_{p,k}\sqrt{V_k}$	$Z_{p,k}^*\sqrt{V_k}$
$K=2$	$m=5$	0.083	0.088	0.063	88.5	88.1	87.3
	10	0.064	0.065	0.053	83.0	83.0	82.4
	15	0.057	0.057	0.050	81.6	81.6	81.1
	20	0.058	0.058	0.053	81.8	81.7	81.4
	50	0.058	0.058	0.056	80.9	80.9	80.8
$K=4$	$m=5$	0.063	0.067	0.054	79.5	79.2	79.0
	10	0.059	0.060	0.053	78.1	78.1	77.9
	15	0.057	0.058	0.053	77.6	77.6	77.5
	20	0.055	0.055	0.052	77.2	77.2	77.1
	50	0.052	0.052	0.050	76.5	76.5	76.4
$K=6$	$m=5$	0.060	0.061	0.052	76.8	76.7	76.6
	10	0.057	0.058	0.054	76.0	76.0	75.9
	15	0.055	0.055	0.052	75.8	75.8	75.7
	20	0.052	0.052	0.051	75.4	75.4	75.4
	50	0.056	0.056	0.055	75.8	75.8	75.8
$K=8$	$m=5$	0.058	0.059	0.053	73.9	73.9	73.8
	10	0.053	0.053	0.050	73.1	73.1	73.0
	15	0.051	0.051	0.049	72.9	72.9	72.9
	20	0.053	0.053	0.051	72.7	72.7	72.7
	50	0.052	0.052	0.051	72.6	72.6	72.6

le rispettivamente, con parametri ignoti. Le osservazioni provenienti dai due campioni hanno uguale numerosità campionaria. Una misura della differenza tra i trattamenti potrebbe essere  $\theta = p_E - p_C$ , ma di solito non viene utilizzata (cfr. Whitehead 1997, Paragrafo 3.2). Si è scelto quindi di utilizzare come parametro di interesse il *log odds ratio*

$$\theta = \log \left\{ \frac{p_E(1 - p_C)}{p_C(1 - p_E)} \right\}$$

In questo modo  $\theta$  può assumere valori nell'intervallo  $(-\infty, \infty)$  e l'approssimazione  $S_k(\theta) \sim N(\theta V_k(\theta), V_k(\theta))$  è solitamente adeguata. Come parametro

**Tabella 4.3:** Suddivisione di  $\hat{\alpha}$  in probabilità di rifiutare l'ipotesi nulla per superiorità e per inferiorità per il test triangolare doppio.

		$S_k$		$Z_{p,k}\sqrt{V_k}$		$Z_{p,k}^*\sqrt{V_k}$	
		$\alpha_{sup}$	$\alpha_{inf}$	$\alpha_{sup}$	$\alpha_{inf}$	$\alpha_{sup}$	$\alpha_{inf}$
$K = 2$	$m = 5$	0.0419	0.0415	0.0440	0.0441	0.0313	0.0315
	10	0.0317	0.0327	0.0321	0.0332	0.0261	0.0274
	15	0.0285	0.0284	0.0287	0.0286	0.0251	0.0250
	20	0.0286	0.0296	0.0288	0.0297	0.0260	0.0269
	50	0.0297	0.0281	0.0297	0.0281	0.0285	0.0272
$K = 4$	$m = 5$	0.0312	0.0320	0.0334	0.0339	0.0267	0.0269
	10	0.029	0.0302	0.0292	0.0305	0.0259	0.0274
	15	0.028	0.0293	0.028	0.0294	0.0259	0.0276
	20	0.0276	0.0271	0.0277	0.0272	0.0263	0.0259
	50	0.0257	0.0259	0.0258	0.0258	0.0253	0.0252
$K = 6$	$m = 5$	0.0302	0.0297	0.0307	0.0302	0.026	0.0257
	10	0.0282	0.0293	0.0282	0.0294	0.0262	0.0275
	15	0.0281	0.0271	0.0282	0.0272	0.0269	0.0256
	20	0.0262	0.0262	0.0263	0.0263	0.0254	0.0255
	50	0.0274	0.0282	0.0274	0.0282	0.0271	0.0278
$K = 2$	$m = 5$	0.0289	0.0293	0.0292	0.0296	0.0260	0.0267
	10	0.0268	0.0265	0.0268	0.0266	0.0253	0.0251
	15	0.0255	0.0257	0.0255	0.0257	0.0244	0.0248
	20	0.0262	0.0266	0.0262	0.0267	0.0254	0.0261
	50	0.0267	0.0251	0.0267	0.0251	0.0264	0.0249



di disturbo si può considerare

$$\lambda = \log\left(\frac{p_E p_C}{(1 - p_E)(1 - p_C)}\right).$$

La log verosimiglianza congiunta dei due gruppi di osservazioni, utilizzando la riparametrizzazione scelta, per  $k = 1, \dots, K$ , è

$$l(\theta, \lambda; x_k, y_k) = \frac{1}{2}(x_k + y_k)\lambda + \frac{1}{2}(y_k - x_k)\theta - mk \log(1 + e^{\frac{\lambda + \theta}{2}}) - mk \log(1 + e^{\frac{\lambda - \theta}{2}}).$$

Le stime di massima verosimiglianza dei parametri sono facilmente ottenibili sfruttando l'equivarianza, con  $\hat{p}_C^{(k)} = x_k/mk$  e  $\hat{p}_E^{(k)} = y_k/mk$ , dove  $x_k$  e  $y_k$  sono le osservazioni campionarie riferite alla  $k$ -esima analisi. La stima del parametro di interesse vincolata a  $\theta = 0$  è  $\hat{\lambda}_0^{(k)} = 2 \log((x_k + y_k)/(2mk - x_k - y_k))$ .

In fase di progettazione dell'esperimento viene calcolata l'informazione di attesa parziale corrispondente alla numerosità campionaria massima dell'esperimento,  $mK$ , ottenendo quindi  $V_{max} = p_C(1 - p_C)(2mK - 1)/4$ . Risolvendo la (4.6) è possibile trovare il valore di  $\delta$  che andrà sostituito nelle (3.1) per determinare le costanti  $a$  e  $c$ . Una volta trovati i valori di  $a$  e  $c$  è possibile tramite la (3.3) trovare il valore di  $V_{max}^c$ . Imponendo  $V_{max}^c = p_C(1 - p_C)(2mK - 1)/4$ , mantenendo inalterato il valore di  $K$  è possibile trovare quale sarà il valore di  $m$  effettivamente utilizzato.

Una volta definite le barriere si può procedere alla conduzione dell'esperimento. Al fine di monitorare l'errore di primo tipo, i dati per entrambi i campioni vengono generati dal modello prescelto sotto l'ipotesi nulla di uguaglianza dei trattamenti. In questo caso, per ogni trattamento, vengono generati  $mK$  dati da una binomiale con probabilità di successo  $p_C = p_E = 0.5$ .

Ogni  $m$  osservazioni viene svolta un'analisi in cui si calcolano  $I_k$ ,  $S_k$  e i

valori delle soglie corrispondenti al valore di  $I_k$  calcolato. In questo modello

$$S_k = \frac{y_k - x_k}{2},$$

$$I_k = \frac{(x_k + y_k)(2mk - x_k - y_k)}{8mk}.$$

Il valore di  $Z_{p,k}$  viene calcolato utilizzando la formula  $Z_{p,k} = \text{sgn}(\hat{\theta}^{(k)})\{2(l_P(\hat{\theta}^{(k)}) - l_P(0))\}^{1/2}$  sostituendo nella log verosimiglianza i valori delle stime di massima verosimiglianza e delle stime vincolate, le cui espressioni sono state prima definite.

Il valore di  $Z_{p,k}^*$  viene calcolato utilizzando una formulazione semplificata (Pace e Salvan, 1996, Esempio 11.6) utilizzabile per le famiglie esponenziali naturali. Se si utilizza la parametrizzazione canonica, infatti, è possibile scrivere la funzione di log verosimiglianza nel seguente modo:

$$l(\theta, \lambda) = \theta t + \lambda u - mkG(\theta, \lambda)$$

dove  $G(\theta, \lambda)$  è una funzione che dipende dal parametro di interesse e da quello di disturbo. In questo caso si ha che  $t = (y_k - x_k)/2$ ,  $u = (y_k + x_k)/2$  e  $G(\theta, \lambda) = \log(1 + e^{\frac{\lambda + \theta}{2}}) - \log(1 + e^{\frac{\lambda - \theta}{2}})$ .

L'espressione per il test  $Z_k^*$  è

$$Z_{p,k}^* = Z_{p,k} + \frac{1}{Z_{p,k}} \log(C_k v_k / Z_{p,k}) \quad (4.7)$$

con

$$C_k = \frac{|G_{\lambda\lambda}(\hat{\theta}^{(k)}, \hat{\lambda}^{(k)})|^{1/2}}{|G_{\lambda\lambda}(0, \hat{\lambda}_0^{(k)})|^{1/2}} \quad \text{e} \quad (4.8)$$

$$v_k = (\hat{\theta}^{(k)} - 0) I_k(\hat{\theta}^{(k)})^{-1/2}$$

dove si ricorda che  $I_k(\hat{\theta}^{(k)})$  è l'informazione osservata parziale per  $\theta$  calcolata in  $(\hat{\theta}^{(k)}, \hat{\lambda}^{(k)})$ , e  $G_{\lambda\lambda}(\cdot, \cdot)$  identifica la derivata seconda di  $G(\cdot, \cdot)$  rispetto a  $\lambda$ . L'informazione  $I_k(\hat{\theta}^{(k)})$  ha la seguente espressione (cfr. Pace e Salvan 1996,

Paragrafo 5.7)

$$I_k(\hat{\theta}^{(k)}) = mk(G_{\theta\theta}(\hat{\theta}^{(k)}, \hat{\lambda}^{(k)}) - G_{\theta\lambda}(\hat{\theta}^{(k)}, \hat{\lambda}^{(k)})[G_{\lambda\lambda}(\hat{\theta}^{(k)}, \hat{\lambda}^{(k)})]^{-1}G_{\lambda\theta}(\hat{\theta}^{(k)}, \hat{\lambda}^{(k)}))$$

Per il modello considerato si ottiene

$$C_k = \sqrt{\frac{(mk - y_k)y_k + (mk - x_k)x_k}{mk(y_k + x_k) - (y_k + x_k)^2/2}}$$

e

$$v_k = \hat{\theta}^{(k)} \frac{1}{\sqrt{mk}} \sqrt{\frac{(mk - y_k)(mk - x_k)x_k y_k}{(mk - y_k)y_k + (mk - x_k)x_k}}$$

I risultati ottenuti dalle simulazioni sono riportati nelle Tabelle 4.4 e 4.5. L'errore di simulazione sul calcolo della probabilità di errore di primo tipo è stimato pari a 0.00097.

Per il calcolo  $Z_{p,k}$  e  $Z_{p,k}^*$ , si è dovuto risolvere il problema della presenza di dati sulla frontiera, che causa stime di massima verosimiglianza pari a  $\pm\infty$ , e quindi rende impossibile il calcolo di  $l(\hat{\theta}, \hat{\lambda})$ . Il problema è stato risolto considerando caso per caso i valori che potevano assumere le osservazioni provenienti dai due gruppi, ossia  $x_k = mk$ ,  $x_k = 0$ , o,  $0 < x_k < mk$  e lo stesso per la variabile casuale  $Y_k$ . In totale si delineano quindi 9 casi possibili. Quando  $0 < x_k < mk$  e  $0 < y_k < mk$  non si hanno problemi. In tutti i casi in cui si è sulla frontiera, si pone  $Z_{p,k}^* = Z_{p,k}$  e il valore di  $Z_{p,k}$  viene ottenuto nel seguente modo. Se  $x_k = 0$  e  $y_k = 0$  oppure  $x_k = mk$  e  $y_k = mk$  si pone  $Z_{p,k} = 0$ . Negli altri casi si nota che la log verosimiglianza profilo tende ad una costante per  $\theta \rightarrow +\infty$  o  $\theta \rightarrow -\infty$ , a seconda dei casi, e si utilizza il valore di tale costante al posto di  $l(\hat{\theta}, \hat{\lambda})$ .

La numerosità campionaria per un test a numerosità campionaria fissata unilaterale, utilizzata per calcolare  $E(mk^*)_{\%}$  è (cfr. Moreno et al., 2002)

$$n_{f,1} = \frac{(\Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta))^2}{\delta^2 p_C (1 - p_C)}$$

dove si ricorda che  $\delta$  è il valore del parametro di interesse sotto l'ipotesi nulla.

Nel caso di test bilaterale,  $n_{f,2}$  si ottiene sostituendo  $\alpha/2$  al posto di  $\alpha$  nella formula di  $n_{f,1}$ .

Confrontando tra loro i tre test, si nota che il test  $Z_{p,k}^* \sqrt{V_k(\theta)}$  necessita di una numerosità campionaria di poco superiore agli altri due test, accentuata soprattutto nel test bilaterale e per  $K = 2, 4$  nel test unilaterale. La numerosità campionaria finale per gli altri due test risulta invece essere molto simile. Per quanto riguarda i valori stimati di  $\alpha$ , non c'è una netta preferenza di un test rispetto ad un altro, tuttavia il migliore pare essere  $S_k$ . Per quanto riguarda il test triangolare singolo, il test  $Z_{p,k}^* \sqrt{V_k(\theta)}$  ha un'approssimazio-

**Tabella 4.4:** Valori di  $\hat{\alpha}$  e numerosità campionaria attesa per i tre tipi di test ottenuti con 50000 simulazioni di un test triangolare, osservazioni generate da  $\text{Bin}(mk, p)$  con  $p = 0.5$ . Parametro di interesse *log odds ratio*.

		Errore di I tipo			$E(mk^*)\%$		
		$S_k$	$Z_{p,k} \sqrt{V_k}$	$Z_{p,k}^* \sqrt{V_k}$	$S_k$	$Z_{p,k} \sqrt{V_k}$	$Z_{p,k}^* \sqrt{V_k}$
$K = 2$	$m = 5$	0.030	0.019	0.020	79.3	79.9	88.7
	10	0.025	0.036	0.024	81.6	80.3	86.3
	15	0.032	0.032	0.028	81.5	81.5	85.2
	20	0.032	0.032	0.036	83.0	83.0	85.6
	50	0.028	0.028	0.029	81.3	81.3	82.6
$K = 4$	$m = 5$	0.022	0.022	0.023	71.7	73.1	74.3
	10	0.020	0.022	0.021	71.6	71.6	71.6
	15	0.027	0.027	0.026	70.1	70.1	70.8
	20	0.027	0.030	0.027	70.8	70.5	71.2
	50	0.025	0.027	0.026	71.2	71.2	70.8
$K = 6$	$m = 5$	0.026	0.033	0.027	71.7	71.7	71.2
	10	0.025	0.027	0.024	69.4	69.4	67.4
	15	0.025	0.026	0.024	68.7	68.7	67.2
	20	0.025	0.024	0.025	69.1	68.6	67.6
	50	0.026	0.024	0.025	68.6	68.6	67.6
$K = 8$	$m = 5$	0.028	0.029	0.027	67.8	67.6	67.2
	10	0.027	0.029	0.028	65.9	65.7	66.4
	15	0.026	0.027	0.026	65.9	65.9	65.7
	20	0.026	0.026	0.027	65.7	65.5	65.7
	50	0.025	0.025	0.026	65.8	65.8	65.6

ne migliore del test  $Z_{p,k}\sqrt{V_k(\theta)}$ , ma questo non viene confermato per il test triangolare doppio. La modifica apportata al test di Whitehead non sembra quindi in questo caso apportare i miglioramenti attesi.

In riferimento al test triangolare doppio sono state stimate le percentuali di errore di primo tipo per superiorità e inferiorità del trattamento sperimentale conteggiate separatamente. La relativa tabella non è stata riportata, ma si ottengono risultati in linea con quelli della Tabella 4.3. Risultati analoghi sono stati ottenuti cambiando il valore del parametro,  $p_C \neq 0.5$ .

**Tabella 4.5:** Valori di  $\hat{\alpha}$  e numerosità campionaria attesa per i tre tipi di test, ottenuti con 50000 simulazioni di un test triangolare doppio, osservazioni generate da  $\text{Bin}(mk, p)$  con  $p = 0.5$ . Parametro di interesse *log odds ratio*.

		Errore di I tipo			$E(mk^* \%)$		
		$S_k$	$Z_{p,k}\sqrt{V_k}$	$Z_{p,k}^*\sqrt{V_k}$	$S_k$	$Z_{p,k}\sqrt{V_k}$	$Z_{p,k}^*\sqrt{V_k}$
$K = 2$	$m = 5$	0.088	0.062	0.062	80.5	81.0	90.6
	10	0.057	0.057	0.062	90.4	90.4	90.1
	15	0.058	0.070	0.063	86.4	85.5	92.8
	20	0.065	0.065	0.075	85.7	85.8	89.2
	50	0.049	0.049	0.053	80.6	80.6	83.0
$K = 4$	$m = 5$	0.053	0.052	0.054	82.3	82.7	86.0
	10	0.051	0.052	0.057	77.7	77.6	79.8
	15	0.057	0.057	0.060	78.6	78.6	80.8
	20	0.051	0.052	0.055	77.0	77.0	78.3
	50	0.053	0.055	0.055	76.6	76.4	77.0
$K = 6$	$m = 5$	0.052	0.066	0.057	78.6	77.8	81.1
	10	0.049	0.053	0.050	76.1	75.9	77.6
	15	0.047	0.052	0.048	75.4	75.1	76.0
	20	0.049	0.050	0.052	75.2	75.1	76.1
	50	0.050	0.050	0.052	75.3	75.3	75.7
$K = 8$	$m = 5$	0.055	0.058	0.061	73.9	73.9	78.3
	10	0.054	0.056	0.056	73.2	73.0	76.0
	15	0.045	0.052	0.047	72.4	71.9	74.1
	20	0.051	0.052	0.053	72.3	72.2	73.8
	50	0.050	0.050	0.052	72.6	72.6	73.6

Si osservi come, in questo esempio, il valore dell'informazione parziale osservata,  $I_k$ , dipenda dai dati, pertanto analisi equispaziate nel numero di osservazioni possono non esserlo nei livelli di informazione. A tal proposito si è pensato di utilizzare una riparametrizzazione tale da rendere il valore di  $I_k$  indipendente dalle osservazioni, per vedere se produca un impatto sulle proprietà dei tre test. Per farlo si è utilizzata la trasformazione arcoseno per una variabile casuale binomiale (Cochran e Cox, 1957).

### 4.5.3 Trasformazione arcoseno per il modello binomiale

Si consideri lo stesso modello del paragrafo precedente, con le stesse assunzioni. Data una variabile casuale con distribuzione  $\text{Bin}(n, p)$ , la trasformazione arcoseno riparametrizza il modello utilizzando l'arcoseno della radice del parametro  $p$ . Si dimostra che (Cochran e Cox, 1957)

$$\arcsin \sqrt{\hat{p}} \sim N(\arcsin \sqrt{p}, \frac{1}{4n}).$$

Tale trasformazione consente quindi di stabilizzare la varianza dello stimatore di massima verosimiglianza. Ai fini degli esperimenti sequenziali comparativi che si stanno considerando, viene utilizzata tale trasformazione per entrambi i modelli, cioè quello per il trattamento sperimentale e per il trattamento di controllo.

Grazie alla trasformazione arcoseno si ottiene che  $I_k = V_k$ , quindi non si ha distorsione prodotta dal fatto che i valori osservati dell'informazione siano diversi da quelli attesi. Viene definito il parametro di interesse  $\theta = \arcsin \sqrt{p_E} - \arcsin \sqrt{p_C}$  e  $\lambda = p_C$ .

La log verosimiglianza della distribuzione congiunta dei due modelli, uti-

lizzando la nuova riparametrizzazione è, per  $k = 1, \dots, K$ ,

$$l(\theta, \lambda; x_k, y_k) = y \log(\sin^2(\theta + \arcsin\sqrt{p_C})) + (mk - y_k) \log(\cos^2(\theta + \arcsin\sqrt{p_C})) \\ + x_k \log p_C + (mk - x_k) \log(1 - p_C).$$

Le stime di massima verosimiglianza dei parametri sono facilmente ottenibili sfruttando l'equivarianza, con  $\hat{p}_C^{(k)} = x_k/mk$  e  $\hat{p}_E^{(k)} = y_k/mk$ , dove  $x_k$  e  $y_k$  sono le osservazioni campionarie fino alla  $k$  esima analisi. La stima del parametro di interesse vincolata a  $\theta = 0$  è  $\hat{\lambda}_0^{(k)} = (x_k + y_k)/(2mk)$ .

L'informazione di parziale attesa calcolata in fase di progettazione dell'esperimento è  $V_{max} = 2mK$  e successivamente vengono calcolati i valori di  $a$  e  $c$  per determinare le barriere utilizzando le (3.1). Le espressioni di  $I_k$  e di  $S_k$  sono

$$I_k = 2mk, \\ S_k = \frac{(y_k - x_k)2mk}{\sqrt{(y_k + x_k)(2mk - y_k - x_k)}}.$$

Il valore di  $Z_{p,k}$  viene ricavato dalla formula  $Z_{p,k} = \text{sgn}(\hat{\theta}^{(k)})\{2(l_P(\hat{\theta}^{(k)}) - l_P(0))\}^{1/2}$  sostituendo nella log verosimiglianza del modello i valori delle stime di massima verosimiglianza e delle stime vincolate, sopra definite. La statistica  $Z_{p,k}^*$  è calcolata utilizzando la (4.4) e la (4.5).

La funzione di log verosimiglianza espressa tramite i parametri e le loro stime di massima verosimiglianza, per  $k = 1, \dots, K$ , è:

$$l(\theta, \lambda; \hat{\theta}^{(k)}, \hat{\lambda}^{(k)}) = 2mk \sin^2(\hat{\theta}^{(k)} + \arcsin\sqrt{\hat{\lambda}^{(k)}}) \log(\sin(\theta + \arcsin\sqrt{\lambda})) + \\ 2mk(1 - \sin^2(\hat{\theta}^{(k)} + \arcsin\sqrt{\hat{\lambda}^{(k)}})) \log(\cos(\theta + \arcsin\sqrt{\lambda})) + \\ mk\hat{\lambda}^{(k)} \log \lambda + mk(1 - \hat{\lambda}^{(k)}) \log(1 - \lambda).$$

La matrice che compare al numeratore di  $U_{p,k}$  ha dimensione  $(2 \times 2)$  e il suo generico elemento verrà indicato con  $[A^{(k)}]_{(i,j)}$ . Per  $k = 1, \dots, K$ , i suoi

elementi sono:

$$\begin{aligned}
[A^{(k)}]_{(1,1)} &= 4mk \sin(\omega) \cos(\omega) [\log(\sin(\omega)) - \log(\cos(\omega)) \\
&\quad - \log(\sin(\nu)) + \log(\cos(\nu))], \\
[A^{(k)}]_{(1,2)} &= \frac{2mk \sin(\omega) \cos(\omega)}{\sqrt{\hat{\lambda}^{(k)} - \hat{\lambda}^{2(k)}}} [\log(\sin(\omega)) - \log(\cos(\omega)) - \log(\sin(\nu)) \\
&\quad + \log(\cos(\nu))] + mk \log(\hat{\lambda}^{(k)}) - mk \log(\lambda_0^{(k)}) + mk \log(1 - \lambda_0^{(k)}) \\
&\quad - mk \log(1 - \hat{\lambda}^{(k)}), \\
[A^{(k)}]_{(2,1)} &= \frac{2mk}{\sqrt{\lambda_0^{(k)} - \lambda_0^{2(k)}}} \sin(\omega) \cos(\omega) [\cot(\nu) + \tan(\nu)], \\
[A^{(k)}]_{(2,2)} &= \frac{mk \sin(\omega) \cos(\omega) [\cot(\nu) + \tan(\nu)]}{\sqrt{\lambda_0^{(k)} \hat{\lambda}^{(k)} (1 - \lambda_0^{(k)}) (1 - \hat{\lambda}^{(k)})}} + \frac{mk}{\lambda_0^{(k)}} + \frac{mk}{1 - \lambda_0^{(k)}},
\end{aligned}$$

dove  $\omega = \hat{\theta}^{(k)} + \arcsin \sqrt{\hat{\lambda}^{(k)}}$  e  $\nu = \arcsin \sqrt{\lambda_0^{(k)}}$ .

Essendo  $\lambda$  scalare,  $j_{\lambda\lambda,k}(\hat{\psi}_\theta)$  è scalare, e la sua formula è

$$\begin{aligned}
j_{\lambda\lambda,k}(\hat{\psi}_0^{(k)}) &= \frac{y(1 + \cot^2(\nu)) + (mk - y_k)(1 + \tan^2(\nu))}{2\lambda_0^{(k)}(1 - \lambda_0^{(k)})} + \frac{mk - x_k}{(1 - \lambda_0^{(k)})^2} + \\
&\quad \frac{[y_k(\cot(\nu)) - (mk - y_k)\tan(\nu)](1 - 2\lambda_0^{(k)})}{2(\lambda_0^{(k)}(1 - \lambda_0^{(k)}))^{3/2}} + \frac{x}{\lambda_0^{2(k)}},
\end{aligned}$$

dove  $\nu$  è definito come sopra. La matrice di informazione  $j_k(\hat{\psi})$ , di dimensione  $2 \times 2$  ha i seguenti elementi:

$$\begin{aligned}
[j_k(\hat{\psi})]_{(1,1)} &= 2[y_k(1 + \cot^2(\omega)) + (mk - y_k)(1 + \tan^2(\omega))], \\
[j_k(\hat{\psi})]_{(1,2)} &= [j_k(\hat{\psi})]_{(2,1)} = \frac{y_k(1 + \cot^2(\omega)) + (mk - y_k)(1 + \tan^2(\omega))}{\sqrt{\hat{\lambda}^{(k)}(1 - \hat{\lambda}^{(k)})}}, \\
[j_k(\hat{\psi})]_{(2,2)} &= \frac{y_k(1 + \cot^2(\omega)) + (mk - y_k)(1 + \tan^2(\omega))}{2\hat{\lambda}^{(k)}(1 - \hat{\lambda}^{(k)})} + \frac{mk - x_k}{(1 - \hat{\lambda}^{(k)})^2} + \\
&\quad \frac{[y_k(\cot(\omega)) - (mk - y_k)(\tan(\omega))](1 - 2\hat{\lambda}^{(k)})}{2(\hat{\lambda}^{(k)}(1 - \hat{\lambda}^{(k)}))^{3/2}} + \frac{x_k}{\hat{\lambda}^{2(k)}},
\end{aligned}$$



dove  $\omega$  è definito come sopra.

I risultati delle simulazioni sono riportati nelle Tabelle 4.6 e 4.7. La numerosità richiesta da un test a numerosità campionaria fissata unilaterale con questa riparametrizzazione, utile per il calcolo di  $E(mk^*)\%$ , è (cfr. Moreno et al., 2002)

$$n_{f,1} = \frac{(\Phi^{-1}(1 - \alpha) + \Phi^{-1}(1 - \beta))^2}{2\delta^2}.$$

Nel caso di test bilaterale è possibile ottenere  $n_{f,2}$  semplicemente sostituendo  $\alpha/2$  al posto di  $\alpha$  nell'espressione di  $n_{f,1}$ . L'errore di simulazione è pari a 0.00097.

**Tabella 4.6:** Valori di  $\hat{\alpha}$  e numerosità campionaria attesa per i tre test ottenuti con 50000 simulazioni di un test triangolare singolo, osservazioni generate da  $\text{Bin}(mk, p)$  con  $p = 0.5$ . Parametro di interesse  $\text{Arcsin}\sqrt{p_E} - \text{Arcsin}\sqrt{p_C}$ .

		Errore di I tipo			$E(mk^*)\%$		
		$S_k$	$Z_{p,k}\sqrt{V_k}$	$Z_{p,k}^*\sqrt{V_k}$	$S_k$	$Z_{p,k}\sqrt{V_k}$	$Z_{p,k}^*\sqrt{V_k}$
$K = 2$	$m = 5$	0.030	0.016	0.012	73.4	74.3	74.3
	10	0.022	0.036	0.020	77.8	76.6	77.9
	15	0.032	0.032	0.024	75.7	75.7	76.3
	20	0.032	0.032	0.032	81.8	81.8	81.8
	50	0.024	0.024	0.023	81.5	81.5	81.6
$K = 4$	$m = 5$	0.026	0.023	0.021	69.6	69.6	69.6
	10	0.020	0.022	0.019	68.5	68.4	68.5
	15	0.026	0.028	0.025	68.2	68.0	68.2
	20	0.027	0.030	0.027	69.2	69.1	69.2
	50	0.026	0.027	0.026	70.8	70.7	70.8
$K = 6$	$m = 5$	0.027	0.033	0.025	70.1	69.3	71.6
	10	0.026	0.027	0.025	67.5	67.4	67.5
	15	0.025	0.027	0.024	67.0	66.8	67.0
	20	0.026	0.026	0.025	67.9	67.2	67.9
	50	0.025	0.025	0.025	67.8	67.8	67.9
$K = 8$	$m = 5$	0.028	0.028	0.027	65.6	65.9	65.9
	10	0.026	0.027	0.025	64.2	63.8	64.0
	15	0.027	0.027	0.027	65.0	65.0	64.9
	20	0.026	0.026	0.026	64.6	64.5	64.7
	50	0.025	0.025	0.025	65.4	65.3	65.4

Eventuali problemi nel calcolo di  $Z_{p,k}$  e quindi di  $Z_{p,k}^*$ , dovute alla presenza di osservazioni con valori sulla frontiera, sono stati risolti inducendo una leggera perturbazione dei dati prendendo a modello la correzione di Firth (1993).

Anche in questo caso, in generale, per tutte e tre le statistiche test, i valori di  $\alpha$  tendono a convergere al loro valore nominale al crescere di  $mK$ . La stima di  $E(mK^*)_{\%}$  decresce al crescere di  $mK$ , e a parità di  $mK$ , decresce al crescere di  $K$ .

Confrontando i tre test si nota che in questo caso la stima di  $E(mK^*)_{\%}$  ha

**Tabella 4.7:** Valori di  $\hat{\alpha}$  e numerosità campionaria attesa per i tre test ottenuti con 50000 simulazioni di un test triangolare doppio, osservazioni generate da  $\text{Bin}(mk, p)$  con  $p = 0.5$ . Parametro di interesse  $\text{Arcsin}\sqrt{p_E} - \text{Arcsin}\sqrt{p_C}$ .

		Errore di I tipo			$E(mk^*)_{\%}$		
		$S_k$	$Z_{p,k}\sqrt{V_k}$	$Z_{p,k}^*\sqrt{V_k}$	$S_k$	$Z_{p,k}\sqrt{V_k}$	$Z_{p,k}^*\sqrt{V_k}$
$K = 2$	$m = 5$	0.060	0.031	0.009	77.5	70.0	74.5
	10	0.044	0.072	0.035	76.9	75.0	76.5
	15	0.064	0.064	0.037	76.2	76.2	79.3
	20	0.064	0.064	0.038	83.9	83.8	77.1
	50	0.048	0.048	0.046	83.4	83.4	80.1
$K = 4$	$m = 5$	0.054	0.047	0.033	78.5	79.1	75.4
	10	0.040	0.044	0.038	73.2	73.1	73.6
	15	0.051	0.056	0.039	73.3	73.1	73.8
	20	0.054	0.060	0.045	76.3	75.8	75.4
	50	0.052	0.053	0.048	76.7	76.5	76.7
$K = 6$	$m = 5$	0.054	0.065	0.038	77.1	76.5	75.2
	10	0.052	0.054	0.042	75.0	74.9	75.0
	15	0.049	0.054	0.044	73.3	73.0	74.1
	20	0.050	0.051	0.046	74.9	74.9	75.1
	50	0.050	0.050	0.050	74.8	74.8	74.9
$K = 8$	$m = 5$	0.056	0.057	0.041	72.4	72.5	70.6
	10	0.053	0.055	0.043	72.4	72.2	72.2
	15	0.056	0.057	0.046	71.6	71.8	72.1
	20	0.053	0.053	0.048	72.6	72.6	72.2
	50	0.051	0.051	0.049	72.4	69.9	72.6

un valore molto simile per tutti e tre i test. Spesso ha un valore leggermente inferiore per la statistica  $Z_{p,k}\sqrt{V_k(\theta)}$ , ma si tratta di una differenza minima. Confrontando i valori di  $\alpha$ , la statistica che sembra essere migliore è ancora  $S_k$  soprattutto per valori piccoli di  $m$  e  $K$ . Anche  $Z_{p,k}^*\sqrt{V_k(\theta)}$  non sembra comportarsi male, ma non è superiore a  $S_k$  come qualità. Per quanto riguarda il test triangolare singolo, il test  $Z_{p,k}^*\sqrt{V_k(\theta)}$  è leggermente migliore del test  $Z_{p,k}\sqrt{V_k(\theta)}$ , ma questo non viene confermato per il test triangolare doppio. Come nell'esempio precedente, il risultato non è cambiato simulando dati con valore del parametro  $p_C \neq 0.5$ . I valori delle percentuali di errore di primo tipo stimate per superiorità e inferiorità del trattamento sperimentale, in riferimento al test triangolare doppio sono molto simili, perciò il test è simmetrico. La relativa tabella non è stata qui riportata, poiché i valori sono conformi a quelli della Tabella 4.3.



# Considerazioni finali

Gli esperimenti sequenziali a gruppi sono caratterizzati dal fatto che i dati vengono analizzati più volte durante lo studio. Le analisi vengono svolte quando si hanno a disposizione un gruppo di osservazioni di numerosità prefissata. Nel contesto degli esperimenti sequenziali a gruppi si utilizza, tra gli altri, il test di Whitehead (1997).

Il lavoro di questa tesi è stato finalizzato a verificare se l'utilizzo di metodi asintotici di verosimiglianza di ordine superiore applicati al test di Whitehead produce dei miglioramenti a livello di accuratezza nel controllo dell'errore di I tipo. Per farlo si è adattata la teoria sui test sequenziali presentata nei Capitoli 2 e 3 alle considerazioni teoriche riportate nei Paragrafi 4.2 e 4.3.

Sono state proposte due varianti al test di Whitehead, che nella sua forma originale utilizza come test la funzione *score*. La prima variante è asintoticamente equivalente ma utilizza come test la versione unilaterale del test log rapporto di verosimiglianza,  $Z_{p,k}$ . La seconda variante utilizza una versione modificata di  $Z_{p,k}$ ,  $Z_{p,k}^*$ , che dovrebbe ridurre l'errore di approssimazione. Per tutte le tre versioni, tramite simulazione sono stati stimati la numerosità campionaria attesa come percentuale di quella di un test a campione fisso e l'errore di primo tipo. Tali quantità sono state prese in considerazione per confrontare le tre statistiche test.

Come applicazioni si sono considerati tre casi: il caso di dati normali con varianza ignota e il caso di dati binomiali con due diversi parametri di interesse. I risultati ottenuti nelle simulazioni confermano quanto atteso solo nel primo caso. Per dati normali con varianza ignota, il test di Whitehead

che utilizza  $Z_{p,k}^*$  risulta più accurato, in particolare per piccole numerosità campionarie; il test che utilizza  $Z_{p,k}$  è invece sostanzialmente equivalente a quello originale.

Negli altri due casi invece il test originale pare essere la scelta migliore, andando quindi in contrasto con quanto atteso.

Una possibile spiegazione della divergenza dei risultati ottenuti da quelli attesi nei due casi sopra menzionati potrebbe essere trovata in Pierce e Bellio (2006). Infatti, la statistica  $Z_{p,k}^*$ , espressa nella forma (4.7), può essere riscritta come

$$Z_{p,k}^* = Z_{p,k} + \text{NP} + \text{INF}$$

dove  $\text{NP} = Z_{p,k}^{-1} \log C_k$  controlla l'aggiustamento di  $Z_{p,k}$  per tener conto della presenza di parametri di disturbo, mentre  $\text{INF} = Z_{p,k}^{-1} \log(v_k/Z_{p,k})$  apporta una correzione per migliorare l'approssimazione normale. L'aggiustamento dovuto ad INF può essere sostanziale solo se l'informazione per il parametro di interesse  $\theta$  è piccola, ma solitamente è piccolo o trascurabile. Nel caso di vari parametri di disturbo l'aggiustamento dovuto a NP può essere notevole anche quando l'informazione per il parametro di interesse  $\theta$  è grande.

Com'è noto, la definizione di una regola d'arresto negli esperimenti sequenziali può essere vista come una specificazione sul modello che non influenza la log verosimiglianza. Al contrario, questa specificazione influenza il comportamento campionario delle quantità legate alla verosimiglianza. In particolare, c'è una dipendenza del secondo ordine nell'aggiustamento dovuto a NP, ma del primo ordine nella componente INF. Tuttavia questo è limitato alle regole d'arresto da loro trattate, e potrebbero esserci dei casi in cui NP è sempre indipendente dalla regola di arresto.

Nella fattispecie dei test sequenziali a gruppi, le osservazioni sono assunte indipendenti, ma il gruppo  $k$ -esimo viene osservato solo condizionatamente al fatto che la regola di arresto non sia stata violata con il  $(k - 1)$ -esimo gruppo. Sarebbe opportuno quindi tener conto di questo tipo di dipendenza, anche nella derivazione di  $Z_{p,k}^*$ .

È noto inoltre che i modelli per dati discreti possono presentare pro-

---

blemi per l'accuratezza dei metodi asintotici. Questo può essere un'altra giustificazione del comportamento osservato nel modello binomiale.

Infine, le conclusioni precedenti sono limitate al test triangolare semplice e triangolare doppio di Whitehead, per gli esempi presentati nel Capitolo 4. Potrebbe valere la pena considerare forme funzionali alternative per le barriere d'arresto nel test di Whitehead o in altri tipi di test, in particolare quelle che portano più frequentemente ad arresti anticipati.





# Appendice A

## Codice delle simulazioni

Di seguito viene riportato il codice utilizzato in R, con il quale si sono ottenuti i risultati presentati nel Capitolo 4. Nel primo esempio, per il test triangolare doppio è riportato l'intero algoritmo, mentre, per il test triangolare singolo è riportata solo la parte che differisce dal test triangolare doppio, ossia la parte del controllo degli sforamenti. Negli ultimi due esempi vengono riportate solo la parte di codice relativa al calcolo delle costanti da utilizzare per derivare le soglie e la parte relativa al calcolo delle statistiche test, poiché il resto è uguale al primo esempio. Nel programma le variabili `S`, `r`, `rstar` indicano i test  $S_k$ ,  $Z_{p,k}$  e  $Z_{p,k}^*$  rispettivamente.

### A.1 Modello normale con varianza ignota

#### Test bilaterale

```
tri_std<-function(sigma, nota="no",alpha=0.05, minput=0, K=0, SIM,
                 seed=FALSE, graf="TRUE")
{
  if (!is.null(seed)) set.seed(seed)
  sigma2=sigma^2
  n_max=K*minput
  V_max=n_max/2
  delta=sqrt(8*log(1/alpha)/V_max)
  print(c("delta",delta))
  a=(2*log(1/alpha)/delta)
  c=delta/4
  diff=V_max/K
}
```

```

nuovoV_max=(sqrt(0.538^2*4/K+(16*a*c))-(0.583*2/sqrt(K)))^2/delta^2
nuovon_max=(2*nuovoV_max)
m=round(nuovon_max/K)
print("V max e mK")
print(c(V_max,n_max))
print("V e mK corretto ")
print(c(nuovoV_max, nuovon_max))
print(c("m corretto", m))

rstar.mat=matrix(0, nrow=SIM, ncol=K)
S.mat=matrix(0, nrow=SIM, ncol=K)
r.mat=matrix(0, nrow=SIM, ncol=K)
V.mat=matrix(0,nrow=SIM, ncol=K)
uu.mat=matrix(0,nrow=SIM, ncol=K)
du.mat=matrix(0,nrow=SIM, ncol=K)
#simulazione
for (sim in 1:SIM)
{
S=V=r=rstar=rep(0,K)
u=d=Up=denUp=numUp=rr=rrstar=rep(0,K)
x=rnorm(m*K, sd=sigma)
y=rnorm(m*K, sd=sigma)
xquad=x^2
yquad=y^2

  for (k in 1:K)
  {
mk=m*k
meanx=mean(x[1:mk])
meany=mean(y[1:mk])
sa2=sum(xquad[1:mk]-meanx^2)/mk
sb2=sum(yquad[1:mk]-meany^2)/mk
mediatot=(meanx+meany)/2
theta0=0
lambda0=(meanx+meany)/2
sigma20=(sa2+sb2)/2+((meanx-meany)/2)^2
sigma0=sqrt(sigma20)
sigma2hat=(sa2+sb2)/2
sigmahat=sqrt(sigma2hat)
thetahat=(meanx-meany)/sigmahat
lambdahat=meany

  if (nota=="no")
  {
V[k]=mk*(0.5-(((meanx-meany)/2)^2/(4*sigma20)))
S[k]=mk*(meanx-meany)/(2*sigma0)
rr[k]=sign(thetahat)*sqrt(2*mk*log(sigma20/sigma2hat))
r[k]=rr[k]*sqrt(V[k])
denUp[k]=sqrt((2*mk^2)/(sigma20^3)*(mk^3/sigma20^3))
numUp[k]=mk^3*sigma2hat*thetahat/sigma20^4
Up[k]=numUp[k]/denUp[k]
rrstar[k]=rr[k]+((1/rr[k])*log(Up[k]/rr[k]))
rstar[k]=rrstar[k]*sqrt(V[k])
}
}

```

```

    if (nota=="si")
    {
      V[k]=mk/2
      S[k]=mk*(meanx-meany)/(2*sigma)
      r[k]=mk*(meanx-meany)/(2*sigma)
    }

    if (k==1)
    {
      uu[k]=a+(c*V[k])-(0.583*sqrt(V[k]))
      du[k]=-a+(3*c*V[k])+(0.583*sqrt(V[k]))
    }
    else
    {
      uu[k]=a+(c*V[k])-(0.583*sqrt(V[k]-V[(k-1)]))
      du[k]=-a+(3*c*V[k])+(0.583*sqrt(V[k]-V[(k-1)]))
    }
    ud[k]=-du[k]
    dd[k]=-uu[k]
  }
S.mat[sim,]=S
r.mat[sim,]=r
rstar.mat[sim,]=rstar
V.mat[sim,]=V
uu.mat[sim,]=uu
du.mat[sim,]=du
ud.mat[sim,]=ud
dd.mat[sim,]=dd
}

#controllo sfioramenti e calcolo degli alpha

sup=S.mat>=uu.mat
inf=S.mat<=dd.mat
mezzo=(S.mat>=ud.mat)&(S.mat<=du.mat)
sup.r=r.mat>=uu.mat
inf.r=r.mat<=dd.mat
mezzo.r=(r.mat>=ud.mat)&(r.mat<=du.mat)
sup.rstar=rstar.mat>=uu.mat
inf.rstar=rstar.mat<=dd.mat
mezzo.rstar=(rstar.mat>=ud.mat)&(rstar.mat<=du.mat)

primosup=apply(sup,1,function(x) min(which(x==TRUE)))
primoinf=apply(inf,1,function(x) min(which(x==TRUE)))
primomezzo=apply(mezzo,1,function(x) min(which(x==TRUE)))
primosup.r=apply(sup.r,1,function(x) min(which(x==TRUE)))
primoinf.r=apply(inf.r,1,function(x) min(which(x==TRUE)))
primomezzo.r=apply(mezzo.r,1,function(x) min(which(x==TRUE)))
primosup.rstar=apply(sup.rstar,1,function(x) min(which(x==TRUE)))
primoinf.rstar=apply(inf.rstar,1,function(x) min(which(x==TRUE)))
primomezzo.rstar=apply(mezzo.rstar,1,function(x) min(which(x==TRUE)))

A=cbind(primosup,primoinf,primomezzo)

```

```

B=cbind(primosup.r,primoinf.r,primomezzo.r)
Bstar=cbind(primosup.rstar,primoinf.rstar,primomezzo.rstar)

fine=apply(A,1,which.min)
fine.r=apply(B,1,which.min)
fine.rstar=apply(Bstar,1,which.min)

#quando per una certa simulazione non si hanno arresti anticipati,
#la riga corrispondente a tale simulazione della matrice "primi"
#riporta in tutti gli elementi il valore Inf. In quei casi il minimo
#di primi restituisce sempre 1 (cioè arresto in favore di H1). Devo
#quindi togliere questi casi da quelli in cui effettivamente il primo
#arresto è in favore di H1

storno=0
storno.r=0
storno.rstar=0
for (sim in 1:SIM)
{
if((A[sim,1]==Inf)&(A[sim,2]==Inf)&(A[sim,3]==Inf)) storno=storno+1
if((B[sim,1]==Inf)&(B[sim,2]==Inf)&(B[sim,3]==Inf)) storno.r=storno.r+1
if((Bstar[sim,1]==Inf)&(Bstar[sim,2]==Inf)&(Bstar[sim,3]==Inf))
    storno.rstar=storno.rstar+1
}

contasup=sum(fine==1)-storno
containf=sum(fine==2)
contamezzo=sum(fine==3)
contasup.r=sum(fine.r==1)-storno.r
containf.r=sum(fine.r==2)
contamezzo.r=sum(fine.r==3)
contasup.rstar=sum(fine.rstar==1)-storno.rstar
containf.rstar=sum(fine.rstar==2)
contamezzo.rstar=sum(fine.rstar==3)

alphatot=(contasup+containf)/SIM
alphatot.r=(contasup.r+containf.r)/SIM
alphatot.rstar=(contasup.rstar+containf.rstar)/SIM
alphasup=contasup/SIM
alphasup.r=contasup.r/SIM
alphasup.rstar=contasup.rstar/SIM
alphainf=containf/SIM
alphainf.r=containf.r/SIM
alphainf.rstar=containf.rstar/SIM
print("alpha per superiorità")
print(c(alphasup, alphasup.r, alphasup.rstar))
print("alpha per inferiorità")
print(c(alphainf, alphainf.r, alphainf.rstar))
print("arresto anticipato per H_0=nessuna differenza")
print(c(contamezzo, contamezzo.r, contamezzo.rstar))

kdistr=apply(A,1,min)
kdistr.r=apply(B,1,min)
kdistr.rstar=apply(Bstar,1,min)

```

```

par(mfrow=c(3,2))
if (!contamezzo==0)
  barplot(table(kdistr),xlab="k",ylab="Freq", main="S")
if (!contamezzo.r==0)
  barplot(table(kdistr.r),xlab="k", ylab="Freq",main="Z_ep")
if (!contamezzo.rstar==0)
  barplot(table(kdistr.rstar),xlab="k", ylab="Freq",main="Z*_ep")

kdistr[kdistr==Inf]<-K
kdistr.r[kdistr.r==Inf]<-K
kdistr.rstar[kdistr.rstar==Inf]<-K
ek=mean(kdistr)
ekr=mean(kdistr.r)
ekrstar=mean(kdistr.rstar)

riga0<-----valore di alpha (per S, r e rstar) ----"
names(riga0)<-" "
print(riga0, quote=F)
print(c(alphatot,alphatot.r, alphatot.rstar))
tab=data.frame(S.mat," ", uu.mat)
nome="valori statistica S, e soglia superiore"
names(nome)<-" "
names(tab)=" "
if (K==2) names(tab)<-c("S_1","S_2"," ","u_1","u_2")
if (K==4) names(tab)<-c("S_1","S_2","S_3","S_4"," ",
  "u_1","u_2","u_3","u_4")
if (K==6) names(tab)<-c("S_1","S_2","S_3","S_4","S_5","S_6"," ",
  "u_1","u_2","u_3","u_4","u_5","u_6")
if (K==8) names(tab)<-c("S_1","S_2","S_3","S_4","S_5","S_6","S_7",
  "S_8"," ","u_1","u_2","u_3","u_4","u_5","u_6","u_7","u_8")

uscita=list(nome,tab,alphatot,alphatot.r,alphatot.rstar,ek,ekr,ekrstar)

if (graf=="TRUE")
{
#creo le soglie Continue :
Inn=seq(0,V_max,0.01)
uuc=a+(c*Inn)
duc=-a+(3*c*Inn)
udc=-duc
ddc=-uuc
#GRAFICO SOGLIE PER S
plot(Inn,uuc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
  xlab="V", ylab="S",type="l")
par(new=T)
plot(Inn,duc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
  xlab="V", ylab="S",type="l")
par(new=T)
plot(Inn,udc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
  xlab="V", ylab="S",type="l")
par(new=T)
plot(Inn,ddc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
  xlab="V", ylab="S",type="l")
}

```

```

#SAMPLE PATH (POSSO DECIDERE RISPETTO A QUALE SIMULAZIONE)
points(V.mat,S.mat)
#points(V.mat[2,],S.mat[2,], col="green")
#GRAFICO SOGLIE PER r
plot(Inn,uuc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
      xlab="V", ylab="r",type="l")
par(new=T)
plot(Inn,duc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
      xlab="V", ylab="r",type="l")
par(new=T)
plot(Inn,udc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
      xlab="V", ylab="r",type="l")
par(new=T)
plot(Inn,ddc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
      xlab="V", ylab="r",type="l")
#SAMPLE PATH (POSSO DECIDERE RISPETTO A QUALE SIMULAZIONE)
points(V.mat,r.mat)
#GRAFICO SOGLIE PER rstar
plot(Inn,uuc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
      xlab="V", ylab="rstar",type="l")
par(new=T)
plot(Inn,duc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
      xlab="V", ylab="rstar",type="l")
par(new=T)
plot(Inn,udc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
      xlab="V", ylab="rstar",type="l")
par(new=T)
plot(Inn,ddc,xlim=c(0,(V_max+10)), ylim=c((min(ddc)-20),(max(uc)+20)),
      xlab="V", ylab="rstar",type="l")
#SAMPLE PATH (POSSO DECIDERE RISPETTO A QUALE SIMULAZIONE)
points(V.mat,rstar.mat)
}
return(uscita)
}
#confronto tra vari K per stat score e r e rstar:

simulazioni=50000
analisi=8
s1=tri_std(alpha=0.05, px=0.5, minput=5, K=analisi, SIM=simulazioni,
            seed=12344,graf="TRUE")
s2=tri_std(alpha=0.05, px=0.5, minput=10, K=analisi, SIM=simulazioni,
            seed=12344,graf="TRUE")
s3=tri_std(alpha=0.05, px=0.5, minput=15, K=analisi, SIM=simulazioni,
            seed=12344,graf="TRUE")
s4=tri_std(alpha=0.05, px=0.5, minput=20, K=analisi, SIM=simulazioni,
            seed=12344,graf="TRUE")
s5=tri_std(alpha=0.05, px=0.5, minput=50, K=analisi, SIM=simulazioni,
            seed=12344,graf="TRUE")

alphascore_k8=c(s1[3],s2[3],s3[3],s4[3],s5[3])
alphar_k8=c(s1[4],s2[4],s3[4],s4[4],s5[4])
alpharstar_k8=c(s1[5],s2[5],s3[5],s4[5],s5[5])
ek8=c(s1[6],s2[6],s3[6],s4[6],s5[6])
ekr8=c(s1[7],s2[7],s3[7],s4[7],s5[7])

```

```

ekrstar8=c(s1[8],s2[8],s3[8],s4[8],s5[8])

par(mfrow=c(1,1))
valm=c(1,2,3,4,5)
plot(valm,alphascore_k8,ylim=c(0.0,0.1),xlab="m", ylab="alpha",
      col="violet", main=paste(c("Triangolare, ALPHA per K="),analisi))
par(new=T)
plot(valm,alphan_k8,ylim=c(0.0,0.1),xlab="m",ylab="alpha", col="blue")
par(new=T)
plot(valm,alphanstar_k8,ylim=c(0.0,0.1),xlab="m",ylab="alpha",col="red")
legend(2,0.45,legend=c("score","r","rstar"), pch=c(1,1,1),
      col=c("violet","blue", "red"))
SE=sqrt(0.05*(1-0.05)/simulazioni)
IC1=0.05+1.96*SE
IC2=0.05-1.96*SE
abline(0.05,0)
abline((IC1),0, lty=2)
abline((IC2),0, lty=2)

```

## Test unilaterale

#controllo sforamenti e calcolo degli alpha

```

sup=S.mat>=uu.mat
arresto=S.mat<=du.mat
sup.r=r.mat>=uu.mat
arresto.r=r.mat<=du.mat
sup.rstar=rstar.mat>=uu.mat
arresto.rstar=rstar.mat<=du.mat

primosup=apply(sup,1,function(x) min(which(x==TRUE)))
primoarresto=apply(arresto,1,function(x) min(which(x==TRUE)))
primosup.r=apply(sup.r,1,function(x) min(which(x==TRUE)))
primoarresto.r=apply(arresto.r,1,function(x) min(which(x==TRUE)))
primosup.rstar=apply(sup.rstar,1,function(x) min(which(x==TRUE)))
primoarresto.rstar=apply(arresto.rstar,1,function(x) min(which(x==TRUE)))

A=cbind(primosup,primoarresto)
B=cbind(primosup.r,primoarresto.r)
Bstar=cbind(primosup.rstar,primoarresto.rstar)

fine=apply(A,1,which.min)
fine.r=apply(B,1,which.min)
fine.rstar=apply(Bstar,1,which.min)

storno=0
storno.r=0
storno.rstar=0
for (sim in 1:SIM)
{
  if((A[sim,1]==Inf)&(A[sim,2]==Inf)) storno=storno+1
  if((B[sim,1]==Inf)&(B[sim,2]==Inf)) storno.r=storno.r+1
  if((Bstar[sim,1]==Inf)&(Bstar[sim,2]==Inf))
    storno.rstar=storno.rstar+1
}

```

```

}

contasup=sum(fine==1)-storno
contarresto=sum(fine==2)
contasup.r=sum(fine.r==1)-storno.r
contarresto.r=sum(fine.r==2)
contasup.rstar=sum(fine.rstar==1)-storno.rstar
contarresto.rstar=sum(fine.rstar==2)

alphatot=contasup/SIM
alphatot.r=contasup.r/SIM
alphatot.rstar=contasup.rstar/SIM

print("alpha per superiorità, S r e rstar")
print(c(alphatot,alphatot.r, alphatot.rstar))
print("arresto anticipato per H_0")
print(c(contarresto, contarresto.r, contarresto.rstar))

```

## A.2 Modello binomiale con parametro di interesse il *log odds ratio*

```

lp<-function(lambda,theta,n,ssx,ssy)
{
  logv=(lambda*(ssx+ssy)/2)+(theta*(ssy-ssx)/2)-(n*log(1+exp((lambda+
    theta)/2)))-(n*log(1+exp((lambda-theta)/2)))
  return(logv)
}
logL<-function(par,data)
{
  n=data[1]
  sx=data[2]
  sy=data[3]
  sy*log(sin(par[1]+asin(sqrt(par[2])))^2)+(n-sy)*log(cos(par[1]+
    asin(sqrt(par[2])))^2)+sx*log(par[2])+(n-sx)*log(1-par[2])
}
logLp=function(theta,data)
{
  -nlminb(0,function(x) -logL(c(theta,x),data),lower=-Inf,
    upper=Inf )$objective
}

tribin<-function(alpha=0.05,px,minput, K=0, SIM,seed=FALSE,graf="TRUE")
{
  if (!is.null(seed)) set.seed(seed)
  n_max=minput*K
  V_max=n_max/2*(px-px^2)
  delta=sqrt(8*log(1/alpha)/V_max)
  a=2/delta*log(1/(alpha))
  c=delta/4
  diff=V_max/K
  nuovoV_max=(sqrt(0.538^2*4/K+(16*a*c))-(0.583*2/sqrt(K)))^2/delta^2
  nuovon_max=(2*nuovoV_max/(px-px^2))

```



```

m=round(nuovon_max/K)
print(c("m",m))
mK=m*K

S=rep(0,K)
rr=rep(0,K)
r=rep(0,K)
rrstar=rep(0,K)
Cp=rep(0,K)
up=rep(0,K)
V=rep(0,K)
rstar=rep(0,K)
Cu=rep(0,K)
uu=rep(0,K)
du=rep(0,K)
ud=rep(0,K)
dd=rep(0,K)
S.mat=matrix(0, nrow=SIM, ncol=K)
r.mat=matrix(0, nrow=SIM, ncol=K)
V.mat=matrix(0,nrow=SIM, ncol=K)
rstar.mat=matrix(0,nrow=SIM, ncol=K)
uu.mat=matrix(0,nrow=SIM, ncol=K)
ud.mat=matrix(0,nrow=SIM, ncol=K)
du.mat=matrix(0,nrow=SIM, ncol=K)
dd.mat=matrix(0,nrow=SIM, ncol=K)

for (sim in 1:SIM)
{
x=rbinom(K,m, px)
y=rbinom(K,m, px)
  for (k in 1:K)
    {
      mk=m*k
      sx=sum(x[0:k])
      sy=sum(y[0:k])
      st=sx+sy
      ft=(mk-sx)+(mk-sy)
      V[k]=(st*ft)/(8*mk)
      S[k]=(sx-sy)/2
      thetahat=log(sy*(mk-sx)/(sx*(mk-sy)))
      lambdahat=log(sx*sy/((mk-sy)*(mk-sx)))
      lambda0=2*log(st/ft)
      theta0=0
#per i dati sulla frontiera:

      if ((sx==0)&(sy==0))
      {
        r[k]=0
        rstar[k]=0
      }

      if ((sx==mk)&(sy==mk))
      {
        r[k]=0

```

```

    rstar[k]=0
  }

  if ((sx>0)&(sx<mk)&(sy==0))
  {
    dati = c(mk,sy,sx)
    differenza=round((logLp(-1000,dati)-logLp(theta0,dati)),5)
    if (!thetahat==0) rr[k]=sign(thetahat)*sqrt(2*(l_hat-l_0))
    r[k]=rr[k]*sqrt(V[k])
    rstar[k]=r[k]
  }

  if ((sy>0)&(sy<mk)&(sx==0))
  {
    dati = c(mk,sy,sx)
    differenza=round((logLp(1000,dati)-logLp(theta0,dati)),5)
    if (!thetahat==0) rr[k]=sign(thetahat)*sqrt(2*(l_hat-l_0))
    r[k]=rr[k]*sqrt(V[k])
    rstar[k]=r[k]
  }

  if ((sy>0)&(sy<mk)&(sx==mk))
  {
    dati = c(mk,sy,sx)
    l_hat=logLp(-100,dati)
    l_0=logLp(theta0,dati)
    differenza=round((l_hat-l_0),5)
    if (!thetahat==0) rr[k]=sign(thetahat)*sqrt(2*(l_hat-l_0))
    r[k]=rr[k]*sqrt(V[k])
    rstar[k]=r[k]
  }

  if ((sx>0)&(sx<mk)&(sy==mk))
  {
    dati = c(mk,sy,sx)
    l_hat=logLp(100,dati)
    l_0=logLp(theta0,dati)
    differenza=round((l_hat-l_0),5)
    if (!thetahat==0) rr[k]=sign(thetahat)*sqrt(2*(l_hat-l_0))
    r[k]=rr[k]*sqrt(V[k])
    rstar[k]=r[k]
  }

  if ((sx==mk)&(sy==0))
  {
    dati = c(mk,sy,sx)
    logLp(-100,dati)
    differenza=round((logLp(-1000,dati)-logLp(theta0,dati)),5)
    if (!thetahat==0) rr[k]=sign(thetahat)*sqrt(2*(l_hat-l_0))
    r[k]=rr[k]*sqrt(V[k])
    rstar[k]=r[k]
  }

  if ((sx==0)&(sy==mk))

```

```

    {
    dati = c(mk, sy, sx)
    differenza=round((logLp(100, dati)-logLp(theta0, dati)), 5)
    if (!thetahat==0) rr[k]=sign(thetahat)*sqrt(2*(l_hat-l_0))
    r[k]=rr[k]*sqrt(V[k])
    rstar[k]=r[k]
    }
#dati non sulla frontiera

if((sx>0)&(sy>0)&(sx<mk)&(sy<mk))
{
l_hat=lp(lambda=lambdahat, theta=thetahat, n=mk, ssx=sx, ssy=sy)
l_0=lp(lambda=lambda0, theta=theta0, n=mk, ssx=sx, ssy=sy)
differenza=round((l_hat-l_0), 3)
if (thetahat==0) rr[k]=sqrt(2*differenza)
if (!thetahat==0) rr[k]=sign(thetahat)*sqrt(2*(l_hat-l_0))
r[k]=rr[k]*sqrt(V[k])

    if(rr[k]==0)
    {
    rrstar[k]=0
    }
    else
    {
    Cp[k]=sqrt(((mk-sy)*sy+(mk-sx)*sx)/(mk*st-(st^2/2)))
    up[k]=thetahat*(1/sqrt(mk))/(sqrt(((mk-sy)*sy+(mk-sx)*sx)))*
    (sqrt((mk-sx)*(mk-sy)*sx*sy))
    Cu[k]=Cp[k]*up[k]
    rrstar[k]=(rr[k]+(1/rr[k])*log(Cu[k]/rr[k]))
    rstar[k]=rrstar[k]*sqrt(V[k])
    }
}

if (k==1)
{
uu[k]=a+(c*V[k])-(0.583*sqrt(V[k]))
du[k]=-a+(3*c*V[k])+(0.583*sqrt(V[k]))
}
else
{
uu[k]=a+(c*V[k])-(0.583*sqrt(V[k]-V[(k-1)]))
du[k]=-a+(3*c*V[k])+(0.583*sqrt(V[k]-V[(k-1)]))
}
ud[k]=-du[k]
dd[k]=-uu[k]
}
S.mat[sim,]=S
r.mat[sim,]=r
V.mat[sim,]=V
rstar.mat[sim,]=rstar
uu.mat[sim,]=uu
du.mat[sim,]=du
ud.mat[sim,]=ud
dd.mat[sim,]=dd

```

```

}

#il resto della funzione è uguale a quella per il primo esempio...
#(...)
}

```

### A.3 Modello binomiale con trasformazione arcoseno

```

cot <- function(x) 1/tan(x)

logL=function(par,data)
{
  n=data[1]
  sx=data[2]
  sy=data[3]
  sy*log(sin(par[1]+asin(sqrt(par[2])))^2)+(n-sy)*log(cos(par[1]+
    asin(sqrt(par[2])))^2)+sx*log(par[2])+(n-sx)*log(1-par[2])
}

tribin<-function(alpha=0.05,px,minput,K=0,SIM,seed=FALSE,graf="TRUE")
{
  if (!is.null(seed)) set.seed(seed)
  n_max=minput*K
  V_max=n_max*2
  delta=sqrt(8*log(1/alpha)/V_max)
  print(c("delta", delta))
  a=2/delta*log(1/(alpha))
  c=delta/4
  nuovoV_max=(sqrt(0.538^2*4/K+(16*a*c))-(0.583*2/sqrt(K)))^2/delta^2
  nuovon_max=(nuovoV_max/2)
  print("V max e mK")
  print(c(V_max,n_max))
  print("nuovo V_max e mK corretto")
  print(c(nuovoV_max,nuovon_max))
  m=round(nuovon_max/K)
  print(c("m",m))
  mK=m*K
  S=rr=r=rrstar=rstar=up=V=rep(0,K)
  numeratore=denominatore=V
  Cu=Cp=rep(0,K)
  uu=du=ud=dd=rep(0,K)
  S.mat=matrix(0, nrow=SIM, ncol=K)
  r.mat=matrix(0, nrow=SIM, ncol=K)
  V.mat=matrix(0,nrow=SIM, ncol=K)
  rstar.mat=matrix(0,nrow=SIM, ncol=K)
  uu.mat=matrix(0,nrow=SIM, ncol=K)
  ud.mat=matrix(0,nrow=SIM, ncol=K)
  du.mat=matrix(0,nrow=SIM, ncol=K)
  dd.mat=matrix(0,nrow=SIM, ncol=K)

  #simulazione

```

```

for (sim in 1:SIM)
{
x=rbinom(K,m, px)
y=rbinom(K,m, px)

  for (k in 1:K)
  {
mk=m*k
sx=sum(x[0:k])
sy=sum(y[0:k])
V[k]=2*mk

    if(!((sx>0)&(sx<mk)&(sy<mk)&(sy>0)))
    {
sy=sy+0.5
sx=sx+0.5
mk=mk+2
}

data=c(mk,sx,sy)
S[k]=2*mk*(sy-sx)/sqrt((sx+sy)*(2*mk-sx-sy))
theta0=0
lambda0=(sy+sx)/(2*mk)
lambdahat=sx/mk
thetahat=asin(sqrt(sy/mk))-asin(sqrt((sx)/(mk)))
cap=thetahat+asin(sqrt(lambdahat))
zero=asin(sqrt(lambda0))
parhat=c(thetahat,lambdahat)
parzero=c(theta0,lambda0)
l_hat=logL(parhat,data)
l_0=logL(parzero,data)
differenza=round((l_hat-l_0),3)
if (!thetahat==0) rr[k]=sign(thetahat)*sqrt(2*(l_hat-l_0))
else rr[k]=0
r[k]=rr[k]*sqrt(V[k])

    if(rr[k]==0)
    {
rrstar[k]=0
}
else
{
theta0=0
num11=4*mk*sin(cap)*cos(cap)*(log(sin(cap))-log(cos(cap))-
  log(sin(zero))+log(cos(zero)))
num12=2*mk*sin(cap)*cos(cap)*(log(sin(cap))-log(cos(cap))-
  log(sin(zero))+log(cos(zero)))/sqrt(lambdahat-lambdahat^2)+
  mk*log(lambdahat)-mk*log(lambda0)+mk*log(1-lambda0)-
  mk*log(1-lambdahat)
num21=2*mk*sin(cap)*cos(cap)*(cot(zero)+tan(zero))/
  sqrt(lambda0-lambda0^2)
num22=mk*sin(cap)*cos(cap)*(cot(zero)+tan(zero))/
  sqrt(lambda0*lambdahat*(1-lambda0)*(1-lambdahat))+
  mk/lambda0+mk/(1-lambda0)
Num=matrix(c(num11,num12,num21,num22),ncol=2,nrow=2,byrow=T)

```

```

numeratore[k]=det(Num)
jlambda=(sy*(1+cot(zero)^2)+(mk-sy)*(1+tan(zero)^2))/
(2*lambda0*(1-lambda0))+(sy*(cot(zero))-(mk-sy)*
(tan(zero)))*(1-2*lambda0)/(2*(lambda0*(1-lambda0))^(3/2))+
sy/lambda0^2+(mk-sy)/(1-lambda0)^2
j11=2*(sy*(1+cot(cap)^2)+(mk-sy)*(1+tan(cap)^2))
j12=(sy*(1+cot(cap)^2)+(mk-sy)*(1+tan(cap)^2))/sqrt(lambdahat
*(1-lambdahat))
j22=(sy*(1+cot(cap)^2)+(mk-sy)*(1+tan(cap)^2))/(2*lambdahat*
(1-lambdahat))+sy*(cot(cap))-(mk-sy)*(tan(cap))*
(1-2*lambdahat)/(2*(lambdahat*(1-lambdahat))^(3/2))+sy/
lambdahat^2+(mk-sy)/(1-lambdahat)^2
Jhat=matrix(c(j11,j12,j12,j22),ncol=2,nrow=2,byrow=T)
J=det(Jhat)
denominatore[k]=sqrt(J)*sqrt(jlambda)
Cu[k]=numeratore[k]/denominatore[k]
rrstar[k]=(rr[k]+(1/rr[k])*log(Cu[k]/rr[k]))
}
rstar[k]=rrstar[k]*sqrt(V[k])

if (k==1)
{
uu[k]=a+(c*V[k])-(0.583*sqrt(V[k]))
du[k]=-a+(3*c*V[k])+(0.583*sqrt(V[k]))
}
else
{
uu[k]=a+(c*V[k])-(0.583*sqrt(V[k]-V[(k-1)]))
du[k]=-a+(3*c*V[k])+(0.583*sqrt(V[k]-V[(k-1)]))
}
ud[k]=-du[k]
dd[k]=-uu[k]
}
S.mat[sim,]=S
r.mat[sim,]=r
V.mat[sim,]=V
rstar.mat[sim,]=rstar
uu.mat[sim,]=uu
du.mat[sim,]=du
ud.mat[sim,]=ud
dd.mat[sim,]=dd
}

#il resto della funzione è uguale a quella per il primo esempio...
#(...)
}

```

# Riferimenti bibliografici

- Armitage, P. e Schneiderman, M. (1958). Statistical problems in a mass screening program. *Ann. New York Academy Sci.*, **76**, 896-908.
- Barnard, G.A. (1946). Sequential tests in industrial statistics. *J. Roy. Statist. Soc. Suppl.*, **8**, 1-26.
- Barndorff-Nielsen, O. E. (1980). Conditionality resolutions. *Biometrika*, **67**, 293-310.
- Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343-365.
- Bartky, W. (1943). Multiple sampling with constant probability. *Ann. Math. Statist.*, **14**, 363-377.
- Cochran, W.G. e Cox, G. M. (1957). *Experimental Designs*. John Wiley and Sons, Inc., New York.
- Dodge, H. F. e Roming, H. G. (1929). A method for sampling inspection. *Bell Syst. Tech. J.*, **8**, 613-631.
- Dunnett, C.W. (1961). The statistical theory of drug screening. In *Quantitative Methods in Pharmacology*, North-Holland, Amsterdam, 212-231.
- Emerson, S.S. e Fleming, T.R. (1989). Symmetric group sequential designs. *Biometrics*, **45**, 905-923.

- Facey, K. M. (1992). A sequential procedure for a phase II efficacy trial in hypercholesterolemia. *Contr. Clin. Trials*, **13**, 122-133.
- Fairbanks, K. e Madsen, R. (1982). P values for tests using a repeated significance test design. *Biometrika*, **69**, 69-74.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27-38.
- Freeman, H.A., Friedman, M., Mosteller, F. e Wallis, W.A. (1948). *Sampling Inspection*. McGraw-Hill, New York.
- Ghosh, B.K. (1991). A brief history of sequential analysis. In *Handbook of Sequential Analysis*. Eds. B.K. Ghosh and P.K. Sen, Marcel Dekker, New York, 1-19.
- Ghosh, B.K. e Sen, P.K. (Eds.) (1991). *Handbook of Sequential Analysis*. Marcel Dekker, New York.
- Jennison, C. e Turnbull, B.W. (1984). Repeated confidence intervals for group sequential clinical trials. *Contr. Clin. Trials*, **5**, 33-45.
- Jennison, C. e Turnbull, B.W. (1989). Interim analyses: the repeated confidence interval approach (with discussion). *J. Roy. Statist. Soc. B.*, **51**, 305-361.
- Jennison, C. e Turnbull, B.W. (1997). Group sequential analysis incorporating covariate information. *J. Amer. Statist. Assoc.*, **92**, 1330-1341.
- Jennison, C. e Turnbull, B.W. (2000). *Group Sequential Methods with application to Clinical Trials*. Chapman & Hall/CRC, Boca Raton.
- Kittelson, J. M. e Emerson, S. S (1999). A unifying family of group sequential test designs. *Biometrics*, textbf55, 874-882.
- Kotz, S. (2006). *Encyclopedia of statistical sciences, Volume 11*. Wiley-Interscience, Cornell University.



- Lai, T. L. (1988). Nearly optimal sequential tests of composite hypothesis. *Ann. Statist.*, **16**, 856-886.
- Lan, K.K.G. e DeMets, D.L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, **70**, 659-663.
- MIL-STD-105E (1989). *Military Standard Sampling Procedures and Tables for Inspection by Attributes*. U.S. Government Printing Office, Washington D.C.
- Moreno, V., Martìn, I., Torres, F., Horas, M., Rios, J. e Gonzalez, J. (2002). Inverse sampling and triangular sequential designs to compare a small porportion with a reference value. *Questiò*, **26**, 259-271.
- O'Brien, P.C. e Fleming, T.R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, **35**, 549-556.
- Pace, L. e Salvan, A. (1996). *Teoria della Statistica*. Cedam, Padova.
- Pace, L. e Salvan, A. (2001). *Introduzione alla Statistica*. Cedam, Padova.
- Pampallona, S. e Tsiatis, A.A. (1994). Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. *J. Statist. Planning and Inference*, **42**, 19-35.
- Pierce, D.A. e Bellio, R. (2006). Effects of the reference set on frequentist inferences. *Biometrika*, **93**, 425-438.
- Pocock, S.J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, **64**, 191-199.
- Proschan, M.A., Follmann, D.A. e Waclawiw, M.A. (1992). Effect of assumption violations on Type I error rate in group sequential monitoring. *Biometrics*, **48**, 1131-1143.

- R Development Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Roseberry, T.D. e Gehan, E.A. (1964). Operating characteristic curves and accept-reject rules for two and three stage screening procedures. *Biometrics*, **20**, 73-84.
- Schneiderman, M. (1961). Statistical problems in the screening search for anticancer drugs by the National Cancer Institute of the United States. In *Quantitative Methods in Pharmacology*, North-Holland, Amsterdam, 232-246.
- Severini, T. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.
- Shao, J. e Feng, H. (2007). Group sequential  $t$ -test for clinical trials with small sample sizes across stages. *Contemp. Clin. Trials*, **28**, 563-571.
- Shih, M-C., Lai, T. L., Heyse, J. F. e Chen, J. (2010). Sequential generalized likelihood ratio tests for vaccine safety evaluation. *Stat. Med.*, **29**, 2698-2708.
- Siegmund, D. (1979). Corrected diffusion approximations in certain random walk problems. *Adv. Appl. Probab.*, **11**, 701-719.
- Siegmund, D. (1985). *Sequential Analysis*. Springer-Verlag, New York.
- Slud, E.V. e Wei, L-J. (1982). Two-sample repeated significance tests based on the modified Wilcoxon statistic. *J. Amer. Statist. Assoc.*, **77**, 862-868.
- Spiegelhalter, D.J., Freedman, L.S. e Parmar, M.K.B. (1994). Bayesian approaches to clinical trials (with discussion). *J. Roy. Statist. Soc. A.*, **157**, 357-416.

- 
- Wald, A. (1945). Sequential tests of statistical hypotheses reviewed. *Ann. Math. Statist.*, **16**, 117-186.
- Wald, A. (1947). *Sequential Analysis*. Wiley, New York.
- Wald, A. e Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.*, **19**, 326-339.
- Wang, S.K. e Tsiatis, A.A. (1987). Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics*, **43**, 193-200.
- Wetherill, G.B. e Glazebrook, K.D. (1986). *Sequential Methods in Statistics*. Chapman and Hall, Londra.
- Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical Trials*. John Wiley & Sons Ltd, Chichester.
- Whitehead, J. e Stratton, I. (1983). Group sequential clinical trials with triangular continuation regions. *Biometrics*, **39**, 227-236.