

**Università degli Studi di Padova**

**Facoltà di Scienze Statistiche**

**Corso di Laurea in Scienze Statistiche, Demografiche e Sociali**



**Tesi di Laurea**

# **Sui Test Multi-Aspetto**

Relatore: Ch.mo Prof. Fortunato Pesarin

Laureando: Aldo Solari

a.a. 2002 / 2003



*A mio fratello*

*Esprimo qui la mia gratitudine a coloro che mi hanno incoraggiato e aiutato nello svolgere questo lavoro, in particolare alla mia famiglia, a Sara, Livio e Luca.*

*Desidero sentitamente ringraziare chi, in questo mio percorso, mi ha fornito i mezzi per iniziarlo, e chi mi ha permesso di portarlo a termine: il prof. Romano Isler e il prof. Fortunato Pesarin.*

*Mi sento ad un tempo debitore e riconoscente nei confronti di particolari amici. Vorrei ringraziarli individualmente, ma sarebbe impossibile riuscirvi senza tralasciare altre persone cui debbo altrettanto. Anche se anonimi, i miei ringraziamenti sono profondamente sentiti.*

# Indice

	<b>PREFAZIONE .....</b>	<b>XI</b>
<b>1</b>	<b>INTRODUZIONE .....</b>	<b>1</b>
<b>2</b>	<b>SULL'ANALISI CONDIZIONATA E NON CONDIZIONATA .....</b>	<b>3</b>
	2.1 Introduzione .....	3
	2.2 Che cosa si può intendere per analisi condizionata? .....	3
	2.3 Un'analisi condizionata non parametrica .....	5
	2.4 Valutazione condizionata e incondizionata .....	5
<b>3</b>	<b>CARATTERIZZAZIONE DEI TEST DI PERMUTAZIONE .....</b>	<b>7</b>
	3.1 Introduzione .....	7
	3.2 Spazio campionario di permutazione associato a $\mathbf{X}$ .....	8
	3.3 Supporto di permutazione indotto da $T$ .....	9
	3.4 Test di permutazione randomizzati e la proprietà di similarità .....	10
	3.5 Test di permutazione non randomizzati .....	10
	3.6 Non distorsione condizionata .....	12
<b>4</b>	<b>ESTENDIBILITÀ DELLE INFERENZE DI PERMUTAZIONE .....</b>	<b>15</b>
	4.1 Introduzione .....	15
	4.2 Uso improprio dei metodi parametrici .....	16
	4.3 Procedure di selezione distorte .....	17
	4.4 Evitare le estensioni inferenziali? .....	19
<b>5</b>	<b>I TEST DI PERMUTAZIONE APPLICATI AGLI STUDI SPERIMENTALI E OSSERVAZIONALI .....</b>	<b>21</b>
	5.1 Introduzione .....	21
	5.2 Studi sperimentali e osservazionali .....	21
	5.3 L'obiettivo del confronto .....	22
	5.4 La comparabilità tra gruppi .....	22
	5.5 La struttura degli studi per effetti causali .....	23
	5.5.1 Un'assunzione critica: assegnazione al trattamento fortemente ignorabile .....	24
	5.5.2 La comparabilità tra gruppi negli studi sperimentali .....	24
	5.5.3 La comparabilità tra gruppi negli studi osservazionali .....	26

5.6	Il controllo dei fattori di confondimento in fase di analisi .....	27
5.6.1	Un'applicazione negli studi osservazionali con stratificazione a posteriori per <i>propensity score</i> .....	28
5.7	Dall'associazione trattamento-risposta alla relazione di tipo causa-effetto .....	30
5.8	I test di ipotesi .....	32
5.8.1	I test di permutazione applicati agli studi sperimentali .....	32
5.8.2	Il test di permutazione di Fisher negli esperimenti randomizzati .....	34
5.8.3	I test di permutazione applicati agli studi osservazionali .....	35
5.8.4	Test di permutazione sotto un modello logistico per <i>propensity score</i> .....	36
5.9	NPC applicata ad uno studio clinico e ad un'indagine epidemiologica .....	38
5.9.1	Studio clinico su un farmaco respiratorio .....	38
5.9.2	Un'indagine epidemiologica: il progetto SETIG .....	39
<b>6</b>	<b>I TEST MULTI-ASPETTO .....</b>	<b>41</b>
6.1	Introduzione .....	41
6.2	Il “ <i>Maximum</i> ” test .....	42
6.3	Approccio Multi-Aspect .....	44
6.4	Procedura generale Multi-Aspect .....	48
6.5	Multi-Aspect come caso particolare della metodologia di combinazione non parametrica per test dipendenti .....	52
6.5.1	Assunzioni sui test parziali .....	52
6.5.2	Proprietà desiderabili delle funzioni di combinazione .....	53
6.5.3	Algoritmo in due fasi per la combinazione nonparametrica .....	54
6.5.4	Non distorsione del test combinato .....	55
6.5.6	Consistenza del test combinato .....	56
6.6	Scelta delle statistiche test parziali di permutazione .....	58
6.7	Scelta delle funzioni di combinazione .....	59
6.8	Esempi in letteratura conformi all'approccio Multi-Aspect .....	60
6.8.1	Analisi statistica di configurazioni spaziali di punti con combinazione di $k$ indicatori .....	60
6.8.2	Casualità e test di casualità nell'analisi crittografica .....	64
6.8.3	Test di adattamento per distribuzioni di cui siano note frequenze e quantità .....	65
6.8.4	Test non parametrici Bi-Aspect per problemi di posizione con applicazione a statistiche mediche, di Business e industriali .....	67

<b>7</b>	<b>POTENZA STATISTICA QUANDO SI ESAMINA LA DIFFERENZIAZIONE</b>	
	<b>GENETICA: UN'APPLICAZIONE DELLA PROCEDURA MULTI-ASPECT .....</b>	<b>71</b>
7.1	Introduzione .....	71
7.2	Presentazione del problema .....	71
7.3	Esemplificazione del problema .....	74
	7.3.1 <i>Summation method</i> .....	75
	7.3.2 Metodo di Bonferroni .....	76
7.4	Conclusioni di Ryman e Jorde .....	76
	7.4.1 <i>Summation method</i> : somma di statistiche Chi-quadrato ...	77
	7.4.2 <i>Summation method</i> : metodo di Fisher .....	79
	7.4.2 Metodo di Bonferroni .....	80
7.5	Soluzione di permutazione .....	81
	7.5.1 Test di permutazione parziali .....	82
	7.5.2 Procedura Multi-Aspect .....	84
7.6	Studio comparativo via simulazione .....	85
7.7	Studio via simulazione quando le differenze riguardano $z \leq k$ loci	90
7.8	Legge di Hardy-Weinberg .....	93
7.9	Studio via simulazione nel caso di dipendenza fra i $k$ loci .....	94
7.10	Controllo della molteplicità .....	97
	7.10.1 Metodi di <i>Closed Testing</i> per test multipli .....	97
	7.10.2 Procedura <i>step-down</i> MinP Bonferroni-Holm .....	98
	7.10.3 Procedura <i>step-down</i> di Tippett .....	100
<b>8</b>	<b>VERIFICA DELL'UGUAGLIANZA DI DUE DISTRIBUZIONI:</b>	
	<b>UN'APPLICAZIONE DELLA PROCEDURA MULTI-ASPECT .....</b>	<b>101</b>
8.1	Introduzione .....	101
8.2	Verifica dell'uguaglianza di due distribuzioni .....	101
	8.2.1 Modello di risposta omoschedastico con effetti additivi fissi .....	102
	8.2.2 Modello di risposta con effetti stocastici generalizzati .....	103
8.3	Procedura Multi-Aspect per la verifica dell'uguaglianza di due distribuzioni .....	104
	8.3.1 Prima fase: definizione degli aspetti .....	104
	8.3.2 Prima fase: scelta delle statistiche test parziali di permutazione .....	105
8.4	Statistiche utilizzate per il <i>distance-aspect</i> .....	106
	8.4.1 Test parziale di permutazione $T_{KS}$ .....	106
	8.4.2 Test parziale di permutazione $T_{AD}$ .....	107
8.5	Statistiche test utilizzate per il <i>location-aspect</i> .....	108

8.5.1	Test parziale di permutazione $T_a$ .....	109
8.5.2	Test parziale di permutazione $T_{Me}$ .....	110
8.6	Test parziale di permutazione $T_s$ per lo <i>scatter-aspect</i> .....	111
8.7	Seconda e terza fase: strutturazione degli aspetti e scelta delle funzioni di combinazione .....	112
8.8	Studio comparativo via simulazione per il modello con effetti additivi fissi .....	115
8.8.1	Modello e ipotesi .....	116
8.8.2	Statistiche test utilizzate .....	117
8.8.3	Distribuzioni di popolazione considerate .....	118
8.9	Analisi del tempo per chiamate di servizio .....	127
8.10	Studio del supporto di permutazione di alcune statistiche test utilizzate .....	129
8.11	Studio via simulazione per il modello con effetti stocastici generalizzati .....	131
8.11.1	Modello e ipotesi .....	131
8.11.2	Statistiche test e distribuzioni considerate .....	133
8.12	Confronto della concentrazione di colesterolo nel plasma dopo aver mangiato a regimi di pesce o di carne .....	136
<b>ALLEGATI .....</b>		<b>139</b>
<b>BIBLIOGRAFIA .....</b>		<b>148</b>



[...] *unsere Untersuchung* [...] *richtet sich nicht auf die Erscheinungen, sondern [...] auf die 'Möglichkeiten' der Erscheinungen.*

[...] *la nostra ricerca* [...] *si rivolge non ai fenomeni, ma [...] alle 'possibilità' dei fenomeni.*

LUDWING WITTGENSTEIN



# Prefazione

I contenuti di questo lavoro si possono sintetizzare come segue:

Nel capitolo 1, si esprimono delle considerazioni in merito agli approcci inferenziali.

Nel capitolo 2, dopo una breve analisi da cui emerge la necessità che nelle inferenze condizionate anche i dati d'osservazione debbano partecipare alla scelta dello stimatore o del test, viene proposta una definizione di analisi inferenziale condizionata in ambito non parametrico. Vengono presentate inoltre le due tipologie di valutazione: quella condizionata e quella incondizionata.

Nel capitolo 3, viene formalmente presentata la teoria dei test di permutazione, che include la definizione dello spazio di permutazione, del supporto di permutazione, dei test randomizzati e non, e infine delle proprietà di similarità e di non distorsione condizionata.

Nel capitolo 4, si considera la situazione che si presenta quando si è interessati ad estendere l'inferenza all'intera popolazione obiettivo; viene messo in evidenza l'uso improprio dei metodi parametrici quando non vengono soddisfatte le assunzioni riguardanti la loro validità, si discute l'attuabilità del concetto di campione casuale e si presentano due esempi di procedure di selezione distorte; infine, si accenna all'estendibilità in senso debole delle inferenze di permutazione in virtù delle proprietà di similarità e di non distorsione condizionata.

Nel capitolo 5, in virtù delle considerazioni del precedente capitolo, vengono analizzati i test di permutazione applicati a due tipi di studi, quello sperimentale e quello osservazionale, con l'obiettivo di confrontare due gruppi caratterizzati da due differenti livelli del fattore di cui si vuole valutare l'effetto. Viene discussa la comparabilità dei gruppi, il controllo dei fattori di confondimento, con particolare riguardo, in fase di analisi, alla tecnica di stratificazione a posteriori per *propensity score* relativa agli studi osservazionali; infine, si esamina il problema dell'interpretazione in senso causale delle relazioni tra il trattamento e la risposta. Alle argomentazioni teoriche e ai vari problemi che insorgono, seguono le soluzioni di permutazione appropriatamente motivate, con special riguardo all'applicazione di test basati sulla combinazione non parametrica di test di permutazione dipendenti (NPC). Si presentano infine due esemplificazioni rispettive ai due tipi di studi presentandone le soluzioni basate su NPC.

Nel capitolo 6, si discute il criterio della verifica multipla, presentandone le motivazioni e gli svantaggi nel caso che essa venga condotta in maniera inappropriata; si presenta inoltre una soluzione descritta in letteratura, il *maximum* test. Dalla considerazione che differenti test sono adatti a verificare differenti aspetti della stessa ipotesi nulla, si presenta l'approccio Multi-Aspetto come caso particolare della metodologia di combinazione non parametrica, con particolare riguardo all'informatività congiunta delle statistiche test, all'impossibilità dell'assunzione di indipendenza dei test parziali e alla difficoltà di

modellarne le relazioni di dipendenza, e quindi della necessità di far fronte ad esse in maniera non parametrica, “catturandole” attraverso la procedura di combinazione. Viene presentata quindi la procedura generale multi-aspect, costituita da tre fasi: (i) definizione degli aspetti e scelta delle statistiche test, (ii) strutturazione degli aspetti, (iii) scelta delle funzioni di combinazione. In relazione a quest’ultima fase, viene considerata una forma di combinazione neutra, denominata procedura di combinazione iterata, che aderisce all’approccio multi-aspect. Si enfatizza inoltre la possibilità di una verifica parziale, o *componetwise testing*. In seguito viene illustrata la metodologia di combinazione non parametrica, introducendo le assunzioni sui test parziali, le proprietà desiderabili delle funzioni di combinazione, e l’algoritmo in due fasi; dopo di che si dimostra la non distorsione e la consistenza del test combinato. In riferimento alla procedura multi-aspect, si forniscono suggerimenti pratici per effettuare una selezione ragionevole dei test parziali e delle funzioni di combinazione. Infine, si presentano quattro esempi dalla letteratura la cui impostazione sottintende un approccio conforme alla logica multi-aspect.

Nel capitolo 7, si esamina un’applicazione della procedura multi-aspect al problema della differenziazione genetica: si presenta uno studio dalla letteratura che valuta la capacità di rilevare eterogeneità genetica combinando le informazioni di più loci, nel quale vengono confrontati due approcci statistici: il *summation method* contro l’applicazione del metodo di Bonferroni. Dopo un’attenta discussione sulle conclusioni di tale studio, viene presentata la soluzione di permutazione conforme alla procedura multi-aspect, con particolare riguardo alla procedura di combinazione iterata, e quindi viene condotto uno studio comparativo via simulazione che replica metodologicamente quello discusso inizialmente. Tale studio viene in seguito esteso ad un caso più generale e realistico, che tiene in considerazione che è a priori non noto se vi sono differenze per quanto riguarda uno, alcuni o tutti i loci. Si illustrano in seguito le condizioni per l’assunzione dell’equilibrio di Hardy-Weinberg e si effettua uno studio di simulazione che prevede una strutturazione delle relazioni di dipendenza dei loci considerati, mettendo in evidenza l’anticonservatività della soluzione ottenuta attraverso il *summation method*. Infine viene enfatizzata l’importanza delle inferenze marginali, presentandone la soluzione attraverso i metodi di *closed testing*.

Nel capitolo 8, si esamina un’ulteriore applicazione della procedura multi-aspect al problema della verifica dell’uguaglianza di due distribuzioni: vengono specificati due modelli di risposta, quello omoschedastico con effetti addittivi fissi, e quello con effetti stocastici generalizzati. Si procede quindi alla definizione degli aspetti e delle statistiche test, alla strutturazione di tali aspetti e alla scelta delle funzioni di combinazione. Si presentano quindi due studi via simulazione, il primo dei quali comparativo ad una soluzione illustrata in letteratura, rispettivamente per il modello con effetti addittivi fissi e per il modello con effetti stocastici generalizzati; per ciascuno dei due studi, vengono inoltre presentate due esemplificazioni pratiche.

Nel lavoro svolto, i capitoli vengono struttati gerarchicamente in sotto-paragrafi; alcune considerazioni vengono presentate come osservazioni, numerate con rispetto al capitolo al quale fanno riferimento. I calcoli *computer-intensive* sono stati effettuati attraverso il *software Matlab*, e i relativi algoritmi verranno presentati in Appendice.

La prassi di riportare integralmente i riferimenti degli Autori può essere giustificata da una citazione di Fisher:

*«If you must write about someone else's work it is, I feel sure, worth taking even more than a little trouble to avoid misrepresenting him. One safeguard is to use actual quotations from his writings; better still a series of comparative quotations»* (lettera indirizzata a Tukey del 27 Aprile 1955).

Consapevole della perfettibilità dell'esito, si è d'accordo con la seguente citazione:

*E renderò pubblico quel poco che ho appreso affinché qualcuno, di me più esperto, possa suggerire il vero, e con la sua opera dimostri e condanni il mio errore. Potrò così rallegrarmi, almeno, di essere stato uno strumento attraverso cui la verità è giunta alla luce.*

ALBRECHT DÜRER

# Capitolo 1

## Introduzione

Questo lavoro tratta i principali aspetti dei metodi di inferenza condizionata basati sull'approccio di permutazione e studia soluzioni appropriate per importanti problemi di interesse pratico.

Partendo dall'opinione di Kempthorne (1955) «*When one considers the whole problem of experimental inference, that is of test of significance, estimation of treatment differences and estimations of errors of estimated differences, there seems little point in present state of knowledge in using a method of inference other than randomization [permutation] analysis*», in accordo con Pesarin (2001) si approva solo parzialmente tale parere, poiché non si è convinti che tutti i problemi inferenziali confluiscono nell'approccio di permutazione.

Si è ritenuto opportuno premettere le seguenti considerazioni di carattere generale:

*«Se volgiamo lo sguardo verso il passato, e consideriamo gli sviluppi delle discipline dentro e intorno alla statistica negli ultimi decenni, potremmo certamente segnalare molte novità di rilievo, scegliendo ciascuno a seconda del proprio orientamento i fatti salienti in disparate direzioni e interpretandoli secondo il proprio punto di vista. Il campo è vasto, gli aspetti fra cui va diviso l'interessamento dei diversi specialisti sono numerosi, ed è sempre vivace il fermento d'idee che scaturisce dai contatti e contrapposizioni fra diverse concezioni, scuole, mentalità. (...)*

*In tesi generale, mi sembra diffusa la tendenza - a mio avviso deplorabile e rovinosa - a spezzettare lo scibile in compartimenti stagni autosufficienti. Può essere ed è spesso utile o necessaria una tale separazione momentaneamente, ai fini espositivi, per fissare le idee lasciando deliberatamente da parte aspetti che allontanerebbero dal punto in questione, per comodità mnemonica, od altro; ma l'utilità sussiste solo a patto di non dimenticare che si tratta di una separazione momentanea tollerabile come tale e non di un'amputazione che si possa pretendere stabile. Nessuna parte del pensiero è viva se avulsa dal tutto, e sembra invece spesso che i cultori di ogni anche minimo campicello si prefiggano come suprema ambizione l'autarchico isolamento» (de Finetti, 1960).*

*«Nella sua essenza il punto di vista che viene sostenuto pone in rilievo l'insufficienza per lo statistico di dover fare riferimento ad una sola impostazione o approccio inferenziale, rispetto alla necessità di dover affrontare i più diversi problemi inferenziali che può incontrare nella sua attività. Di conseguenza viene sostenuta la necessità di dover fare ricorso, nelle appropriate circostanze e per le appropriate finalità, anche a una pluralità di diversi approcci (...)*

*I quali approcci, dovendo rispondere a propri sistemi di assiomi, ovviamente tra loro non tutti coincidenti ed eventualmente solo implicitamente stabiliti attraverso autonome indagini critiche di idoneità e pertinenza (la critica razionale), sono individualmente idonei a rispondere a qualcuno dei desiderata od obiettivi cui uno statistico potrebbe essere interessato a rispondere.*

*Va comunque sottolineato che vi possono essere, e vi sono di fatto, degli approcci inferenziali in grado di rispondere, pur non altrettanto bene gli uni rispetto agli altri, a sottoinsiemi di obiettivi coincidenti ed essere, quindi, tra loro parzialmente in competizione. Sembrerebbe ovvio, in tali casi, di dover dare la preferenza al migliore tra essi limitatamente ai desiderata comuni.*

*Ma non è sempre ovvio stabilire ragionevoli criteri di ordinamento rispetto ai quali valutare quale debba essere il migliore» (Pesarin, 1986).*

*«La visione unitaria, cui ritengo e sostengo sia necessario arrivare, non va intesa come una cosa nuova da costruire dal nulla, bensì come il risultato della riunione, senza alcuna forzatura, in un quadro d'insieme generale, delle diverse teorie parziali esistenti (oggi, domani, o non importa quando)» (de Finetti, 1977).*

# Capitolo 2

## Sull'analisi condizionata e non condizionata

### 2.1 INTRODUZIONE

E' possibile definire la nozione di analisi inferenziale condizionata in molti modi diversi. Si considerano ora alcuni tra i più importanti.

L'idea di analisi condizionata nell'ambito della verifica delle ipotesi illustrata da Lehmann (1986) e anche in Cox e Hinkley (1974) si riferisce al condizionamento a particolari esperimenti (statistiche ancillari) che complessivamente sono descritti da una mistura, in cui ogni componente della mistura descrive un possibile esperimento.

In ambito bayesiano, Raiffa e Schlaifer (1961) propongono un'analisi ad "esperimento attuato", sostituendo l'evento certo,  $\Omega$ , con il campione osservato  $\underline{x}_n = (x_1, \dots, x_n)$ , che è basata sulla nozione di rischio terminale, supponendo l'addittività delle perdite. Un tale approccio coincide in realtà con il calcolo dei punti di ottimo non condizionato come è dettagliatamente descritto nel lavoro di Pesarin e Celant (2000).

E' bene evidenziare la necessità di distinguere tra gli "strumenti" che descrivono in termini probabilistici un condizionamento da quelli che effettivamente permettono un'inferenza condizionata. E' infatti chiaro che sì la densità condizionata, la posteriori, ecc. sono effettivi condizionamenti, ma è altrettanto vero che non sono di per sé necessari o sufficienti per evidenziare l'apporto di informazione definita tramite la realizzazione dell'osservazione o dell'esperimento. Ciò che ci si propone di illustrare in questo capitolo sono alcune nuove nozioni di condizionamento descritte nel lavoro di Pesarin e Celant (2001) che tengano conto del tipo di situazioni che il dato e le informazioni di cui si dispone congiuntamente consentono di descrivere.

### 2.2 CHE COSA SI PUÒ INTENDERE PER ANALISI CONDIZIONATA?

Per definire l'analisi condizionata, si cercherà di capire ciò che caratterizza una analisi non condizionata. A questo fine si consideri un problema di stima parametrica in un'ottica non decisionale e non bayesiana. Come sarà chiaro in seguito le conclusioni a cui si perverrà resteranno valide anche se il contesto è bayesiano, decisionale, ecc.

Nella stima parametrica si ipotizza ("ipotesi di lavoro") di disporre di un modello statistico parametrico  $(\Omega, \beta_{\Omega^n}, \{P_{\theta} : \theta \in \Theta\})$ , che descrive tutti gli esiti di un esperimento. Quindi si



associa al modello un insieme di funzioni (gli stimatori),

$$\tau = \{T : \Omega^n \rightarrow \Theta, (x_1, \dots, x_n) \rightarrow T(x_1, \dots, x_n) = \hat{\theta}_T\}$$

Il problema della stima consiste allora, fissato un criterio di ottimalità  $K$ , nel determinare in  $\tau$  una particolare funzione  $K$ -ottimale,  $T^* : \Omega^n \rightarrow \Theta$

*Osservazione 2.1:* Un caso particolare ancorché consueto si ha quando  $\tau$  è l’insieme degli stimatori non distorti e  $K$  è la varianza (campionaria) dello stimatore. In questo caso  $T^*$  è la soluzione del problema di ottimo,  $T^* : \text{Var}(T^*) = \min_{T \in \tau} \text{Var}(T)$ .

A questo punto, determinato  $T^*$  “si realizza” l’esperimento e si prende come stima di  $\theta_0$  (“vero” valore del parametro) l’immagine  $\hat{\theta}_T = T^*(x_1, \dots, x_n)$  che la mappa  $T^*$  assume sulle osservazioni  $\underline{x}_n = (x_1, \dots, x_n)$ . E’ fondamentale rilevare che in questo ragionamento la conoscenza o meno dell’esito dell’esperimento, e cioè di  $\underline{x}_n$ , sia del tutto irrilevante per definire  $T^*$ . Ciò significa che tutta l’informazione su  $\theta_0$  è essenzialmente contenuta nel modello  $(\Omega, \beta_{\Omega^n}, \{P_\theta : \theta \in \Theta\})$ . Infatti, qualunque sia il risultato  $\underline{x}_n$  dell’esperimento,  $T^*$  non cambia. Questa completa indipendenza di  $T^*$  da  $\underline{x}_n$  è per l’appunto ciò che si ritiene qualificanti (caratterizzi) l’analisi dei dati come una analisi non condizionata, dove diventa ora chiaro che per non condizionata si deve intendere non condizionata all’esito dell’esperimento.

*Osservazione 2.2:* E’ ovvio che  $\underline{x}_n$  sia indispensabile per attribuire una allocazione a  $\theta_0$ . Ma ciò che importa osservare è che la precisione della stima, e quindi tutte le affermazioni probabilistiche che facciamo su  $\theta_0$  (in altri termini, l’inferenza), dipendano solo ed esclusivamente da  $T^*$ . Si ricordi che la specificazione di  $T^*$ , pur non essendo una funzione di  $\theta$ , dipende da tutto  $\Omega^n$ , da  $\{P_\theta : \theta \in \Theta\}$ , da  $\tau$ , da  $K$  e in nessun modo da  $\underline{x}_n$ . A ben pensarci è abbastanza intuitivo che tutta l’informazione necessaria per definire  $T^*$  sia contenuta nel modello  $(\Omega, \beta_{\Omega^n}, \{P_\theta : \theta \in \Theta\})$ . Si pensi, ad esempio, al caso seguente:

$$(\mathfrak{R}^n, \beta_{\mathfrak{R}^n}, \{N(\mu, 1) : \mu \in \mathfrak{R}^n\})$$

L’unica differenza tra  $N(\mu_1, 1)$  e  $N(\mu_2, 1)$  è  $\mu_1 \neq \mu_2$ .  $N(\mu_1, 1), N(\mu_2, 1)$  hanno la stessa variabilità che è ineliminabile al finito ed è indipendente dal “vero” valore  $\mu_0$ .

La precisione dello stimatore dipenderà dalla variabilità dei dati che si osservano e cioè dalla

variabilità,  $\sigma^2$ , di  $N(\mu_1, 1)$ . E questo significa che la precisione di  $T^*$  e dunque la sua specificazione è indipendente da  $\mu$ . D’altra parte, a seconda “della risposta dell’esperimento” e cioè dei dati  $(x_1, \dots, x_n)$ , si stima  $\mu_0$ . E’ chiaro infatti che se  $\mu_0 = 100$  è “più facile” che i dati osservati appartengano all’intervallo  $[100 - \varepsilon, 100 + \varepsilon]$  che non ad esempio all’intervallo  $[1 - \varepsilon, 1 + \varepsilon]$ . Ossia  $T^*$  è definito da  $(\Omega, \beta_{\Omega^n}, \{P_\theta : \theta \in \Theta\})$  ma la stima  $\hat{\theta}$  da  $T^*(\underline{x}_n)$ .

Intuitivamente, per quanto affermato sulla caratterizzazione dell’analisi non condizionata, si potrà allora dire che in un’analisi condizionata le osservazioni  $\underline{x}_n$  non dovranno giocare un ruolo meramente “passivo”. Dovranno cioè portare nuova informazione rispetto a quella contenuta nelle ipotesi di lavoro (il modello statistico). Ossia la stessa definizione dello stimatore, e non solo della stima, dovrà dipendere in qualche modo anche da  $\underline{x}_n$ .

### 2.3 UN’ANALISI CONDIZIONATA NON PARAMETRICA

Nella statistica non parametrica il modello non è specificato in modo rigido come nel caso parametrico. Ciò ci fa intuire che sarà possibile dare una definizione in sintonia con l’idea iniziale che i dati partecipino alla stessa definizione del modello. In questo contesto l’ipotesi di lavoro che si ritiene di assumere è che la distribuzione che genera i dati sia invariante per permutazione (dati scambiabili). In tale caso lo spazio campionario, dato  $\underline{x}_n$ , sarà quello delle permutazioni del campione osservato sul quale consideriamo la distribuzione indotta dalle permutazioni che, in problemi di verifica d’ipotesi e sotto  $H_0$ , è quella uniforme. Ossia il modello di riferimento è il seguente:

$$(\pi(\Omega_{oss}), I(\pi(\Omega_{oss})), U(1, \dots, n!))$$

dove  $\pi(\Omega_{oss})$  è l’insieme delle permutazioni ottenute permutando i dati,  $I(\pi(\Omega_{oss}))$  è l’insieme di tutti i sottoinsiemi di  $\pi(\Omega_{oss})$  ed infine  $U(1, \dots, n!)$  è la distribuzione uniforme definita su  $\pi(\Omega_{oss})$ .

### 2.4 VALUTAZIONE CONDIZIONATA E INCONDIZIONATA

Vi sono circostanze nelle quali le procedure non condizionate di verifica parametrica possono essere appropriate dal punto di vista dell’interpretazione dei relativi risultati inferenziali, così vi sono altre nelle quali possono diventare improprie o anche impossibili da risolvere. Per

contro, vi sono circostanze nelle quali le procedure di verifica condizionata possono diventare appropriate e a volte inevitabili.

Si può inoltre decidere di adottare inferenze di verifica condizionate, non solo quando la controparte non condizionata non sia possibile, ma anche quando si vuole dare più importanza all'insieme di dati osservato rispetto al modello di popolazione.

Infatti le inferenze condizionate possono essere di interesse quando, per qualsiasi ragione, si desidera limitarsi a metodi condizionati esplicitamente e positivamente circoscritti all'effettivo insieme dei dati.

Lehmann (1986) pone la seguente domanda: «*Is a conditional or unconditional point of view more appropriate?*». La risposta, ovviamente, dipende dal contesto. Lehmann suggerisce che se un esperimento verrà effettuato molte volte, ad esempio in un contesto industriale o agricolo, le prestazioni medie possono essere la caratteristica principale di interesse ed il metodo incondizionato adatto. Tuttavia, se le ripetizioni si riferiscono a clienti differenti, o siano potenziali piuttosto che reali, l'interesse si concentrerà sull'evento particolare alla mano, e quindi condizionare può sembrare maggiormente appropriato. Vediamo di capire meglio queste considerazioni attraverso il seguente esempio (Pesarin, 2001).

Nel valutare l'affidabilità delle automobili, un proprietario può essere interessato alla propria macchina, o parco macchine se ne possiede più d'una, in quanto, tenendo in considerazione come esse vengono usate, è responsabile di tutti i costi di affidabilità relativi, dando quindi origine ad una *valutazione condizionata*. Diversamente, il punto di vista del fabbricante di automobili, la cui reputazione e i costi di garanzia sono relativi a tutto l'insieme di macchine simili già prodotte, può essere per lo più centrato su una sorta di comportamento medio, dando origine a una forma di *valutazione non condizionata* relativa all'intera popolazione di automobili.

Entrambi i punti di vista sono importanti e utili nei problemi reali, poiché vi sono situazioni (come quella del proprietario di automobili) nelle quali si può essere interessati all'inferenza condizionata, mentre ve ne sono delle altre (come quella del fabbricante di macchine) nelle quali si può essere interessati all'inferenza non condizionata. Quindi, così come entrambi i punti di vista sono di pratico interesse, entrambi i tipi di inferenza sono di importanza metodologica e spesso possono essere analizzati usando lo stesso insieme di dati.

Comunque, è da sottolineare che, nelle procedure di verifica condizionate, a condizione che la scambiabilità dei dati rispetto ai gruppi venga soddisfatta sotto l'ipotesi nulla, i metodi di permutazione ricoprono un ruolo centrale.

# Capitolo 3

## Caratterizzazione dei test di permutazione

### 3.1 INTRODUZIONE

«Permutation tests are conditional inferential procedures in which conditioning is with respect to the sub-space associated with the set of sufficient statistics in the null hypothesis for all nuisance entities, including the underlying, known or unknown, distribution» (Edgington, 1995; Good, 2000; Pesarin, 2001).

Per chiarire questo concetto, senza perdita di generalità, si consideri un problema a due campioni: sia  $(X, \Omega, P)$  il modello statistico associato al problema,  $f_P$  la densità di  $P$  rispetto a una misura dominante  $\xi$ ,  $\mathbf{X}_j = \{X_{ji}, i = 1, \dots, n_j\}$  l'insieme dei dati con  $n_j$  elementi nel  $j$ -esimo campione o gruppo ( $j = 1, 2$ ) e  $n = n_1 + n_2$  la numerosità campionaria totale. Si noti che non è stata fatta alcuna assunzione riguardante la procedura di campionamento per la raccolta dei dati, pertanto  $P$  rappresenta o la distribuzione della sotto-popolazione selezionata oppure quella dell'intera popolazione. Si supponga che le ipotesi da verificare siano:

$$H_0 : \{X_1 \stackrel{d}{=} X_2\} = \{P_1 = P_2\} \quad \text{contro} \quad H_1 : \{X_1 < \neq > X_2\}$$

dove la notazione ' $< \neq >$ ' indica una relazione di dominanza stocastica (' $<$ ' o ' $>$ ') oppure di disuguaglianza (' $\neq$ ').

La diretta applicazione della definizione mostra che un insieme di statistiche sufficienti per  $P$  sotto  $H_0$  (si noti che  $P$  è la medesima per i due gruppi) è l'insieme congiunto dei dati osservati  $\mathbf{X} = \{X_{ji}, i = 1, \dots, n_j; j = 1, 2\}$ .

Si noti che la scambiabilità dei dati fra i due gruppi, implicata dall'uguaglianza delle due distribuzioni stabilita sotto  $H_0$ , è una condizione necessaria affinché i test di permutazione siano correttamente applicabili. Quindi, per quanto segue, quando si farà riferimento a un test di permutazione, si assumerà implicitamente che la condizione di scambiabilità sia soddisfatta sotto  $H_0$ .

### 3.2 SPAZIO CAMPIONARIO DI PERMUTAZIONE ASSOCIATO A $\mathbf{X}$

Per procedere condizionatamente si deve caratterizzare il sotto-spazio di riferimento  $\Omega_{/\mathbf{X}}^n$  associato all'insieme di statistiche sufficienti sotto l'ipotesi nulla. Esso è definito come l'insieme dei punti  $n$ -dimensionali di  $\Omega^n$  che contengono la stessa informazione su  $P$  contenuta in  $\mathbf{X}$  sotto l'assunto che  $H_0$  sia vera. In altri termini, viene definito come l'insieme dei punti equivalenti con rispetto alla verosimiglianza sottostante. Quindi un tale sotto-spazio di riferimento, o *coset*, deve contenere tutti i punti  $\mathbf{X}^*$  tali che il rapporto di verosimiglianza  $f_P^{(n)}(\mathbf{X})/f_P^{(n)}(\mathbf{X}^*)$  sia indipendente dalla sottostante e possibilmente non nota  $f_P$ .

Si osservi che quando la densità  $f_P^{(n)}$  è scambiabile rispetto ai suoi argomenti, cioè:  $f_P^{(n)}(x_1, \dots, x_n) = f_P^{(n)}(x_{u_1^*}, \dots, x_{u_n^*})$ , in cui  $(u_1^*, \dots, u_n^*)$  è una generica permutazione di  $(1, \dots, n)$ , allora il *coset*  $\Omega_{/\mathbf{X}}^n$  contiene tutte le permutazioni  $\mathbf{X}^*$  dell'insieme congiunto dei dati  $\mathbf{X}$ .

In particolare, quando i dati sono ottenuti attraverso un campionamento casuale da  $P$ , così che  $f_P^{(n)}(\mathbf{X}) = \prod_{1 \leq i \leq n} f_P(X_i)$ , la proprietà di scambiabilità viene naturalmente soddisfatta.

Viene anche soddisfatta quando si adotta il criterio della randomizzazione, che consiste nell'assegnazione casuale delle unità ai livelli di trattamento, e questo anche se tali unità e i relativi dati sono stati ottenuti attraverso procedure distorte.

Solitamente il sotto-spazio di riferimento  $\Omega_{/\mathbf{X}}^n$  viene chiamato *spazio campionario di permutazione associato a  $\mathbf{X}$* . Si noti che se la numerosità campionaria  $n$  è finita, il numero delle permutazioni di  $\mathbf{X}$  risulta finito. Quindi, per ogni  $\mathbf{X}' \in \Omega_{/\mathbf{X}}^n$  la probabilità condizionata  $\Pr\{\mathbf{X}^* = \mathbf{X}' \mid \Omega_{/\mathbf{X}}^n\}$  è sempre definita e risulta

$$\frac{\sum_{\mathbf{X}^* = \mathbf{X}'} f_P^{(n)}(\mathbf{X}^*) \cdot d\xi^n}{\sum_{\mathbf{X}^* \in \Omega_{/\mathbf{X}}^n} f_P^{(n)}(\mathbf{X}^*) \cdot d\xi^n} = \frac{\#[\mathbf{X}^* = \mathbf{X}', \mathbf{X}^* \in \Omega_{/\mathbf{X}}^n]}{\#[\mathbf{X}^* \in \Omega_{/\mathbf{X}}^n]}$$

che è indipendente da  $f_P$ . Di conseguenza, l'inferenza condizionata è *distribution-free* e non parametrica, ed inoltre non viene influenzata dalla procedura di campionamento adottata per raccogliere i dati. In particolare, quando in  $\Omega_{/\mathbf{X}}^n$  vi è solo un punto tale che  $\mathbf{X}^* = \mathbf{X}'$ , cioè quando non vi ripetizioni<sup>1</sup> nell'insieme dei dati, allora tale probabilità è pari a  $1/n!$ . Da ciò risulta che sotto  $H_0$  la distribuzione di probabilità  $\Pr\{\mathbf{X}^* = \mathbf{X}' \mid \Omega_{/\mathbf{X}}^n\}$  è

---

<sup>1</sup> ties;

uniforme su  $\Omega_{/X}^n$  a prescindere dalla distribuzione sottostante  $P$ . Ovviamente, nell'ipotesi alternativa, quando  $P_1 \neq P_2$  per un'insieme di punti con probabilità maggiore di zero, la statistica sufficiente è la coppia di insiemi di dati  $(\mathbf{X}_1, \mathbf{X}_2)$ . Ne consegue che la scambiabilità dei dati tra i due gruppi non è permessa sotto  $H_1$ . Pertanto, quando si è capaci di trovare statistiche che sono sensibili a questa diversità di distribuzioni, si è anche capaci di costruire test di permutazione adatti. Il problema di stabilire quale, fra le molte possibilità, sia la soluzione migliore per il caso in questione, rimane aperto (Lehmann, 1986; Pesarin, 2001).

### 3.3 SUPPORTO DI PERMUTAZIONE INDOTTO DA $T$

Si supponga che  $T : \Omega^n \rightarrow \mathfrak{R}^1$  sia una statistica test appropriata per la quale, senza perdita di generalità, si assuma che valori grandi sono significativi. Si definisca come *supporto di permutazione indotto da  $T$  su  $\mathbf{X}$*  l'insieme  $\tau(\mathbf{X}) = \{T^* = T(\mathbf{X}^*) : \mathbf{X}^* \in \Omega_{/X}^n\}$  contenente tutti i possibili valori assunti da  $T$  quando  $\mathbf{X}^*$  varia nello spazio campionario  $\Omega_{/X}^n$ . Si ipotizzi che  $H_0$  sia vera e si dispongano le  $M = \#\{\mathbf{X}^* \in \Omega_{/X}^n\}$  componenti di  $\tau(\mathbf{X})$  in ordine non decrescente, cioè  $T_{(1)}^* \leq T_{(2)}^* \leq \dots \leq T_{(M)}^*$ , quindi per ogni valore di  $\alpha \in (0, 1)$ ,  $T_\alpha(\mathbf{X}) = T_\alpha = T_{(M_\alpha)}^*$  definisce il valore critico di permutazione associato alla coppia  $(T, \mathbf{X})$ , dove  $M_\alpha = \lfloor (1 - \alpha) \cdot M \rfloor$ , cioè la parte intera di  $(1 - \alpha) \cdot M$ , rappresenta il numero dei valori di permutazione  $T^*$  che sono strettamente inferiori al valore critico  $T_\alpha$ . Si noti che, quando  $T$  è una funzione continua ed  $X$  una variabile continua, per le quali le *ties* hanno probabilità zero, gli elementi di  $\tau(\mathbf{X})$  sono quasi sicuramente distinti, nel senso che con probabilità uno non vi sono punti multipli. Quindi, essi possono essere disposti in ordine crescente:  $T_{(1)}^* < T_{(2)}^* < \dots < T_{(M)}^*$ . Si osservi inoltre che i valori critici di permutazione  $T_\alpha$  dipendono da  $\Omega_{/X}^n$  e non soltanto da  $\mathbf{X}$ ; infatti la relazione  $T_\alpha = T_\alpha(\mathbf{X}) = T_\alpha(\mathbf{X}')$  è soddisfatta per tutti gli  $\mathbf{X}' \in \Omega_{/X}^n$ , poiché i *coset* associati a  $\mathbf{X}$  e  $\mathbf{X}'$  sono sempre coincidenti:  $\Omega_{/X}^n = \Omega_{/X'}^n$ . Questo implica che, per ogni  $\alpha \in (0, 1)$  fissato,  $T_\alpha$  è un valore fissato nel supporto di permutazione  $\tau(\mathbf{X})$ , che varia come  $\mathbf{X}$  varia in  $\Omega^n$ . Questo ed il fatto che in pratica non si conosce se  $H_0$  sia vera o meno dà origine a ovvie difficoltà nell'esprimere le funzioni di potenza condizionate e non condizionate di  $T$  in una forma chiusa.

### 3.4 TEST DI PERMUTAZIONE RANDOMIZZATI E LA PROPRIETÀ DI SIMILARITÀ

La versione randomizzata del test di permutazione  $\phi_R$  associato alla coppia  $(T, \mathbf{X})$  è definita come

$$\phi_R = \begin{cases} 1 & \text{se } T_o > T_\alpha \\ \gamma & \text{se } T_o = T_\alpha \\ 0 & \text{se } T_o < T_\alpha \end{cases}$$

dove  $T_o = T(\mathbf{X})$  è il valore di  $T$  calcolato sui dati osservati  $\mathbf{X}$  e la probabilità di rifiuto  $\gamma$ , quando  $T_o = T_\alpha$  è data da

$$\gamma = [\alpha - \Pr\{T_o > T_\alpha \mid \Omega_{/\mathbf{X}}^n\} / \Pr\{T_o = T_\alpha \mid \Omega_{/\mathbf{X}}^n\}].$$

E' immediato dimostrare che, per tutti gli insiemi di dati  $\mathbf{X} \in \Omega^n$ , il valore atteso condizionato sotto  $H_0$  di  $\phi_R$  soddisfi

$$E\{\phi_R(\mathbf{X}) \mid \Omega_{/\mathbf{X}}^n\} = \int_{\Omega_{/\mathbf{X}}^n} \phi_R(\mathbf{X}) \cdot dF(t \mid \mathbf{X}) = \alpha$$

dove  $F(t \mid \mathbf{X}) = \Pr\{T^* \leq t \mid \Omega_{/\mathbf{X}}^n\}$  rappresenta la distribuzione di permutazione di  $T$  dato  $\mathbf{X}$ , poiché  $T_o$  è distribuito uniformemente su  $\tau(\mathbf{X})$ . Ne consegue che i test di permutazione randomizzati sono forniti della cosiddetta proprietà di *similarità forte*, nel senso che *per tutte le distribuzioni sottostanti  $P$  e per tutti gli insiemi di dati  $\mathbf{X} \in \Omega^n$ , la probabilità di rifiuto condizionata è  $\mathbf{X}$ -invariante sotto  $H_0$ , e questo indipendentemente da come i dati  $\mathbf{X}$  sono stati selezionati da  $\Omega^n$ .*

### 3.5 TEST DI PERMUTAZIONE NON RANDOMIZZATI

Nei contesti applicativi viene generalmente preferita la cosiddetta versione semplice o non randomizzata, definita come

$$\phi_R = \begin{cases} 1 & \text{se } T_o \geq T_\alpha \\ 0 & \text{se } T_o < T_\alpha \end{cases}$$

in cui il tasso di errore di I tipo associato, il cosiddetto  $\alpha$ -valore raggiungibile, sotto  $H_0$  risulta

$$E\{\phi(\mathbf{X}) \mid \Omega_{/\mathbf{X}}^n\} = \alpha_a \geq \alpha$$

Data una qualunque coppia  $(T, \mathbf{X})$ , l' $\alpha$ -valore raggiungibile appartiene all'insieme  $\Lambda_{\mathbf{X}} = \{L(t \mid \mathbf{X}) : dL(t \mid \mathbf{X}) > 0\}$  di punti-gradino della funzione del livello di significatività  $L(t \mid \mathbf{X}) = \Pr\{T^* \geq t \mid \Omega_{/\mathbf{X}}^n\}$ .

$\Lambda_{\mathbf{X}}$  è sempre un insieme discreto, i cui elementi dipendono da  $n$  e  $(T, \mathbf{X})$ . Pertanto, per i test di permutazione non randomizzati, non tutti i valori dei tassi di errore di I tipo sono possibili. Vale la pena notare che, specialmente quando la variabile risposta  $X$  è discreta o quando ci sono *ties* in  $\mathbf{X}$ , gli  $\alpha$ -valori raggiungibili dipendono dall'effettivo insieme di dati  $\mathbf{X}$ , quindi l'insieme  $\Lambda_{\mathbf{X}}$  non è  $\mathbf{X}$ -invariante. Ma quando in  $\tau(\mathbf{X})$  non vi sono punti multipli, allora  $\Lambda_{\mathbf{X}} = \Lambda = \{h/M, h=1, \dots, M\}$  e quindi è  $\mathbf{X}$ -invariante, ma dipende da  $n$  e da  $T$ . In questo caso gli  $\alpha_a$  raggiungibili hanno salti costanti di  $1/M$ . Naturalmente, quando  $n$  non è troppo piccolo e se in  $\tau(\mathbf{X})$  non ci sono punti con molteplicità maggiore di uno (di solito ciò avviene quando in  $\mathbf{X}$  non vi sono valori uguali), in pratica non vi sono differenze fra  $\phi_R$  e  $\phi$ .

Si dovrebbe notare che, se si vuole verificare con un tasso di errore di I tipo desiderato  $\alpha_d$  e si sceglie  $\alpha_d \geq \alpha_a \in \Lambda_{\mathbf{X}}$ , allora i test non randomizzati risultano conservativi. Chiaramente, se l' $\alpha$ -valore desiderato  $\alpha_d \in \Lambda_{\mathbf{X}}$ , si avrà  $\alpha_d = \alpha_a$ . D'ora in poi ci si riferirà alla versione non randomizzata  $\phi$ , mentre il tasso di errore di I tipo raggiungibile verrà indicato con l'usuale simbolo  $\alpha$ .

Si sa che determinare in pratica i valori critici  $T_{\alpha}$  di un test di permutazione può presentare ovvie difficoltà, specialmente quando nella situazione considerata non si conosce se  $H_0$  sia vera o meno. Quindi si è soliti fare riferimento al cosiddetto  $p$ -value associato a  $(T, \mathbf{X})$ . Questo viene definito come

$$\lambda = \lambda(\mathbf{X}) = \lambda(T_o) = L(T_o \mid \mathbf{X}) = \Pr\{T^* \geq T_o \mid \Omega_{/\mathbf{X}}^n\}$$

la cui determinazione può essere ottenuta esattamente attraverso l'enumerazione completa di  $\tau(\mathbf{X})$  oppure stimata, al desiderato grado di precisione, attraverso il metodo *conditional Monte Carlo* (CMC)<sup>2</sup>.

<sup>2</sup> si veda il capitolo 6, paragrafo 6.5.3;



Si noti che  $\lambda$  è funzione non crescente di  $T_0$  e che risulta in relazione uno-a-uno con l' $\alpha$ -valore raggiungibile di  $\phi$ , nel senso che  $\lambda(\mathbf{X}) > \alpha$  implica  $T_0 < T_\alpha$  e vice versa. Comunque, vale la pena notare che per la determinazione pratica di  $\lambda$  non abbiamo bisogno di conoscere se  $H_0$  sia vera o meno.

Se  $X$  è una variabile continua e non degenere e  $T$  una funzione continua, così che gli  $\alpha$ -valori raggiungibili appartengono a  $\Lambda$  per quasi tutti gli insiemi di dati  $\mathbf{X}$ , allora il test di permutazione non randomizzato è fornito della proprietà di similarità nella forma quasi certa. Questo significa che, *gli  $\alpha$ -valori raggiungibili per  $\phi$  sono  $\mathbf{X}$ -invarianti per quasi tutti gli  $\mathbf{X} \in \Omega^n$ .*

Per variabili discrete o miste, nelle quali le *ties* hanno probabilità positiva, gli  $\alpha$ -valori raggiungibili appartengono a  $\Lambda_{\mathbf{X}}$  che dipende dai dati osservati  $\mathbf{X}$ ; di conseguenza i test di permutazione randomizzati perdono la proprietà di similarità per  $n$  finito. Comunque sia, la similarità viene sempre soddisfatta asintoticamente, pertanto si può dire che viene approssimativamente mantenuta in tutte le situazioni.

Una conseguenza immediata della proprietà di similarità, dal momento che viene soddisfatta almeno approssimativamente per quasi tutti gli insiemi di dati e indipendentemente dalla procedura di selezione, è che *il test di permutazione  $\phi$  raggiunge l' $\alpha$ -size sia rispetto alla popolazione selezionata sia rispetto alla popolazione obiettivo, anche quando entrambe sono vagamente definite.*

### 3.6 NON DISTORSIONE CONDIZIONATA

Per semplicità e senza perdita di generalità, si assuma che sotto  $H_1$  la distribuzione di  $X_2$  sia spostata di una quantità  $\delta > 0$  rispetto alla distribuzione di  $X_1$ , così che

$$H_1 : \{X_1 \stackrel{d}{=} (X_2 - \delta)\}.$$

Quindi, si possono rappresentare gli insiemi di dati con la notazione

$$\mathbf{X}(\delta) = \{X_{ji} = \mu + \delta_j + Z_{ji}, i = 1, \dots, n_j; j = 1, 2\},$$

nella quale  $\mu$  è una quantità di disturbo,  $Z_{ji}$  sono errori scambiabili e centrati con distribuzione comune  $P$ ,  $\delta_1 = 0$  e  $\delta_2 = \delta$ . Per rappresentare la struttura dei dati sotto  $H_0$ , si utilizzerà la notazione  $\mathbf{X}(0)$ ; in questo modo e con riferimento ad alternative

unidirezionali, le ipotesi possono essere scritte in modo equivalente come  $H_0 : \{\delta = 0\}$  e  $H_1 : \{\delta > 0\}$ .

Si consideri la statistica

$$T = \sum_{i=1}^{n_2} X_{2i},$$

permutazionalmente equivalente (Pesarin, 2001) al più comune confronto standardizzato di due medie campionarie. Quindi, con chiaro significato dei simboli, le quantità

$$T_o(0) = n_2\mu + \sum_{i=1}^{n_2} Z_{2i} \quad \text{e} \quad T_o(\delta) = n_2(\mu + \delta) + \sum_{i=1}^{n_2} Z_{2i}$$

rappresentano rispettivamente i valori osservati di  $T$  sotto  $H_0$  e  $H_1$ . In modo corrispondente, se  $v^*$  indica il numero di elementi che sono stati scambiati tra i due gruppi in una generica permutazione, si avrà che  $T^*(0) = n_2\mu + \sum_{i=1}^{n_2} Z_{2i}^*$  e  $T^*(\delta) = n_2\mu + (n_2 - v^*)\delta + \sum_{i=1}^{n_2} Z_{2i}^*$  rappresentano i valori di  $T$  nella permutazione considerata. Quindi si avrà

$$\begin{aligned} \lambda(\mathbf{X}(\delta)) &= \Pr\{T^*(\delta) \geq T_o(\delta) \mid \Omega_{\mathbf{X}(\delta)}^n\} \\ &= \Pr\{T^*(0) - v^*\delta \geq T_o(0) \mid \Omega_{\mathbf{X}(0)}^n\} \\ &\leq \Pr\{T^*(0) \geq T_o(0) \mid \Omega_{\mathbf{X}(0)}^n\} = \lambda(\mathbf{X}(0)) \end{aligned}$$

dalla quale si osserva che i  $p$ -values sotto  $H_1$  sono stocasticamente più piccoli di quelli sotto  $H_0$  e questo per qualunque distribuzione sottostante  $P$  e per tutti gli insiemi di dati  $\mathbf{X} \in \Omega^n$ , indipendentemente da come sono stati raccolti.

Questo risultato ci permette di concludere che, *per ogni*  $\alpha \in \Lambda_{\mathbf{X}(0)}$ , *per ogni distribuzione*  $P$  *e per tutti i possibili insiemi di dati*  $\mathbf{X} \in \Omega^n$ , *la statistica*  $T$  *dà origine a un test di permutazione condizionatamente non distorto*, cioè

$$\Pr\{\lambda(\mathbf{X}(\delta)) \leq \alpha \mid \Omega_{\mathbf{X}(\delta)}^n\} \geq \alpha$$

In aggiunta, se  $0 < \delta < \delta'$  allora  $\lambda(\mathbf{X}(\delta')) \leq \lambda(\mathbf{X}(\delta)) \leq \lambda(\mathbf{X}(0))$ . Quindi, con rispetto all'effetto  $\delta$ , i  $p$ -values sono quantità stocasticamente ordinate non-decrescenti, e di conseguenza

$$\Pr\{\lambda(\mathbf{X}(\delta')) \leq \alpha \mid \Omega_{/\mathbf{X}(\delta')}^n\} \geq \Pr\{\lambda(\mathbf{X}(\delta)) \leq \alpha \mid \Omega_{/\mathbf{X}(\delta)}^n\} \geq \alpha.$$

In Pesarin (2001), la proprietà di non distorsione condizionata è generalizzata a problemi multi-campioni e a qualunque forma di dominanza stocastica, incluse quelle associate a effetti casuali in accordo a un modello di risposta come

$$X_{ji} = \mu + \Delta_{ji} + Z_{ji}, \quad i = 1, \dots, n_j; j = 1, 2,$$

in cui  $\Delta_{1i} = 0$  e  $\Delta_{2i} = \delta_2 + W_{2i} > 0$ , dove  $\delta_2$  è l'effetto medio e le deviazioni dagli effetti  $W_{2i}$  possono dipendere da  $(\mu, Z_{2i})$ .

# Capitolo 4

## Estendibilità delle inferenze di permutazione

### 4.1 INTRODUZIONE

In questo capitolo ci si riferirà a problemi di verifica d'ipotesi, giacché questo è il contesto più comune in cui il confronto fra le inferenze condizionate ed incondizionate viene sviluppato in letteratura.

Come struttura generale, si assuma che vi sia una popolazione obiettivo i cui individui sono elementi dello spazio campionario  $\Omega$ . Si assuma inoltre che una variabile d'interesse  $X$  è osservabile su ogni elemento di  $\Omega$  e che associata a  $(X, \Omega)$  esiste una distribuzione di popolazione  $P$ . La denotazione  $(X, \Omega, P)$  sintetizza il modello statistico associato al problema.

Come si è visto nei precedenti capitoli, nell'inferenza non condizionata il tipico spazio di riferimento è l'intero spazio campionario  $\Omega$ , mentre nell'inferenza condizionata lo spazio di riferimento è ottenuto considerando la restrizione di  $\Omega$  al sotto-spazio associato al condizionamento dell'evento d'interesse. Di solito questo sottoinsieme viene indicato con  $\Omega_{/A}$  e contiene i punti di  $\Omega$  che possiedono una data proprietà  $A$ , cioè il *coset* di  $A$ . Naturalmente, si assume che associata ad  $(X, \Omega_{/A})$  vi sia una distribuzione condizionata  $P_A$ . Se si è essenzialmente interessati a questo sottoinsieme, ovvero quando non si è interessati ad estendere l'inferenza al di fuori di esso (come è tipico delle situazioni nelle quali non si desidera estendere l'inferenza all'intera popolazione  $\Omega$ ), non sorge alcun problema. In questo caso è sufficiente che, con rispetto a  $\Omega_{/A}$ , i metodi di inferenza adottati (solitamente statistiche test) possiedano le proprietà gradite come la non-distorsione, la consistenza, l'ottimalità locale, e via dicendo.

Ma quando si è interessati ad estendere l'inferenza all'intera popolazione obiettivo, e, a causa dell'insufficienza di assunzioni, non è concesso l'uso di metodi basati su tutto  $\Omega$ , allora ci si trova di fronte a problemi difficili o addirittura impossibili da risolvere, specialmente nella struttura dell'approccio parametrico.

Per esempio, è noto che, quando ci sono parametri di disturbo da eliminare, per ottenere un'inferenza corretta per l'intera popolazione obiettivo, è necessario, sotto l'ipotesi nulla, che i relativi modelli di verosimiglianza posseggano *boundedly complete statistics*, o

almeno *invariant statistics* per i parametri di disturbo (Cox e Hinkley, 1974; Lehmann, 1986). Ciò implica che i dati devono essere ottenuti attraverso tecniche di campionamento ben pianificate su modelli di verosimiglianza ben specificati, ma è anche noto che in pratica questa situazione non sempre si verifica.

## 4.2 USO IMPROPRIO DEI METODI PARAMETRICI

Ludbrook and Dudley (1998) hanno rilevato che di 252 studi riportati in cinque quotate riviste biomediche, solo in 11 casi (4%) essi sono basati su procedure di campionamento casuale dalle popolazioni date. Tutti gli altri, la maggioranza dei quali riguarda studi clinici randomizzati<sup>1</sup>, sono basati su dati provenienti da metodi di campionamento distorti e in cui gli individui vengono assegnati casualmente ai trattamenti. Tuttavia, i metodi statistici usati, in 8 casi sono correttamente basati sul concetto di campionamento casuale (i.i.d.) da una popolazione normale, in 189 (96%) casi le analisi utilizzano metodi “*t* o *F*” anche se le condizioni per la loro applicabilità non sono soddisfatte, in 38 casi le analisi sono basate sui ranghi e solo in un caso su metodi di permutazione (16 casi non riguardano la verifica d’ipotesi).

Quindi, in circa l’80% delle situazioni (189/236), vengono usati metodi incapaci di garantire le proprietà inferenziali richieste. Gli autori pensano che questo inconveniente metodologico, almeno in parte, possa esser dovuto alle raccomandazioni della maggior parte dei *referee* delle riviste biomediche che non conoscono il ruolo dell’inferenza condizionata associato ai metodi di permutazione, così che «*is much easier to get a manuscript published if one stays with classical tests under the population model*», sebbene questi test vengano applicati impropriamente.

de Finetti (1975) battezza scherzosamente “Robotolatria” la tendenza di chi, anziché cercare di adeguarsi caso per caso alla natura del problema e delle decisioni che, a seconda delle conclusioni, saranno da prendere, preferisce impiegare procedure, modelli e criteri standardizzati – *adhoceries* come li ha definiti I. J. Good – cioè schemi statistici *prêt-à-porter*, con distribuzioni di qualche tipo standard e ipotizzando spesso gratuitamente l’indipendenza stocastica, eccetera.

Si sostiene quindi che «*lo statistico non può sfuggire alla responsabilità che ha di comprendere i procedimenti che egli applica e raccomanda*» (Fisher, 1935).

---

<sup>1</sup> *randomized clinical trials* (RTC);

### 4.3 PROCEDURE DI SELEZIONE DISTORTE

Si è visto nel precedente paragrafo che una procedura di campionamento ben progettata è abbastanza rara nella maggior parte degli studi sperimentali. Per esempio, se si desidera studiare gli effetti di un farmaco sui ratti, le unità da trattare non saranno scelte casualmente dalla popolazione di tutti i ratti, ma saranno selezionate in qualche modo fra quelle disponibili in un laboratorio ed a caso verranno assegnate ai livelli stabiliti di trattamento. Lo stesso si verificherà nella maggior parte degli studi clinici, in cui alcuni dei pazienti presenti in un ospedale verranno assegnati a caso ad uno dei livelli prestabiliti del trattamento.

Per quanto riguarda il contesto osservazionale, nella maggior parte degli studi le unità di osservazione verranno scelte sulla base ad esempio della disponibilità e la partecipazione volontaria delle unità preposte al reclutamento dei pazienti (per esempio, ospedali, medici, centri di cura). Inoltre, il contesto osservazionale è caratterizzato in generale da una propensione minore e più incerta a introdurre assunzioni e ipotesi sulla popolazione di provenienza del campione.

Ne consegue che *«the deliberate use of sampled populations differing from the target populations is likely to remain a standard practice»* (Cochran, 1965).

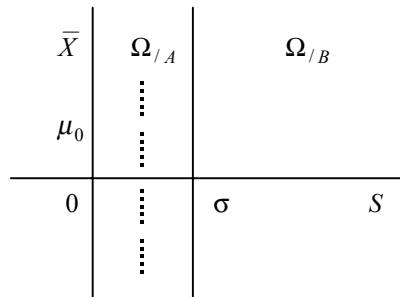
Dunque, si è osservato che il concetto del campione casuale viene attuato raramente nelle applicazioni reali, perché, per vari motivi, il campione effettivo è ottenuto comunemente da procedure di selezione distorte. Ciò implica che la maggior parte delle forme di inferenza incondizionate associate solitamente ai test parametrici, essendo basate sul concetto di campione casuale, sono raramente applicabili nelle situazioni reali.

In generale, quindi, per i test parametrici, l'estensione delle conclusioni inferenziali dal condizionato all'incondizionato può essere fatta solo se l'insieme dei dati è ottenuto attraverso procedure di campionamento ben progettate e applicata all'intera popolazione obiettivo. Questo perché, se i dati sono stati ottenuti da procedure distorte, allora le inferenze associate sono condizionatamente distorte, oppure le stime dei parametri di disturbo non sono *boundedly complete*, o entrambe le cose. Quindi, la similarità o la non distorsione condizionata, oppure entrambe, non vengono assicurate e di conseguenza non vi è la garanzia per il controllo delle proprietà inferenziali richieste.

Per illustrare quest'ultimo aspetto, fra le infinite possibilità di ottenere procedure distorte, si considerino due situazioni che possono essere comuni sia negli studi sperimentali sia in quelli osservazionali (Pesarin, 2001).

La prima considera che, in situazioni con un campione e con errori normalmente distribuiti, lo spazio campionario  $\Omega$  venga suddiviso in due sottoinsiemi di punti  $\Omega_{/A}$  e  $\Omega_{/B}$  definiti rispettivamente da stime “piccole” e “grandi” della varianza non nota, di modo che  $\Omega = \Omega_{/A} \cup \Omega_{/B}$ .

Inoltre si ipotizzi che la specifica procedura distorta consideri i punti campionari di  $\Omega_{/A}$  e che si consegua la coppia di stime  $(\bar{X}_A, S_A)$ . La figura 4.1 descrive tale situazione:



**Figura 4.1:**  
(Pesarin, 2002)

Poiché  $S_A$  è stocasticamente più piccola della corrispondere stima  $S$  che si otterrebbe attraverso il campionamento casuale dall'intera popolazione (notare che in questo contesto  $\bar{X}_A$  ha la stessa distribuzione marginale di  $\bar{X}$ ), la statistica di Student  $t_A = (\bar{X}_A - \mu_0) \cdot \sqrt{n}/S_A$  essendo stocasticamente più grande della corrispondere  $t = (\bar{X} - \mu_0) \cdot \sqrt{n}/S$ , risulta anticonservativa.

Infatti, il tasso di errore di I tipo soddisfa la relazione  $\alpha \leq \alpha_A \leq 1$ . Quindi, in queste situazioni il test  $t$  di Student non può essere utilizzato poiché le proprietà d'invarianza che di *boundedly complete* non vengono soddisfatte.

La seconda situazione considera un caso speciale di distribuzioni cosiddette “pesate” (si veda Patil e Rao, 1977), nel quale

$$f_p(x) \cdot \varphi(x) = K_\varphi \cdot \exp\{-(x - \mu_\varphi)^2 / 2\sigma_\varphi\}$$

dove  $f_p(x)$  è la densità corrispondente alla popolazione obiettivo  $P$ ,  $\varphi(x)$  è una funzione peso non negativa,  $K_\varphi$  è una costante di normalizzazione, ed infine  $\mu_\varphi$  e  $\sigma_\varphi$  sono rispettivamente la media e la deviazione standard della popolazione selezionata  $P_\varphi$ . Quindi, la  $t$  di Student è un test appropriato per  $P_\varphi$ , ma i relativi risultati inferenziali non possono essere estesi alla popolazione obiettivo  $P$ , a meno che la funzione peso sia ben specificata e i corrispondenti parametri ben stimati.

Si noti che, nel contesto del campionamento distorto, può accadere che la media e la varianza campionaria siano stocasticamente indipendenti anche se la popolazione obiettivo non è normale, come nella seconda situazione.

Quindi, in generale, le statistiche basate su procedure distorte possono violare parti importanti della maggior parte dei teoremi basati sul campionamento casuale da una distribuzione di popolazione. Di conseguenza, quando gli esperimenti campionari non sono ben progettati e altrettanto ben condotti, per alcune estensioni le inferenze

parametriche incondizionate diventano improprie e perciò non possono essere usate per gli scopi inferenziali.

#### 4.4 EVITARE LE ESTENSIONI INFERENZIALI?

Il problema dell'uso improprio dei metodi parametrici basati su campionamento i.i.d. dalla popolazione quando vengono usate procedure distorte per raccogliere i dati, suggerisce di limitare sempre le inferenze solo all'effettivo insieme dei dati usando metodi strettamente condizionati e di evitare le estensioni inferenziali all'intera popolazione obiettivo e anche a quella indotta dalla procedura di selezione.

*Osservazione 4.1:* Edgington (1995) afferma che la generalizzazione dei risultati conseguiti con l'utilizzo di un test di permutazione su gruppi sperimentali ottenuti attraverso la randomizzazione di campioni non-casuali dovrebbe essere fatta con cura. In questo caso, a suo parere, l'inferenza dovrebbe limitarsi solo ai soggetti coinvolti nell'esperimento, mentre l'inferenza su altri soggetti dovrebbe essere di tipo non statistico, cioè senza una base probabilistica. Egli sottolinea inoltre che le inferenze non statistiche rappresentano una procedura scientifica molto comune secondo la quale i risultati ottenuti dalle unità osservate vengono generalizzati a tutte le unità che sono simili riguardo a quelle caratteristiche considerate come importanti per la progettazione dell'esperimento.

Comunque, a fini pratici, metodologici e teorici, si è detto che queste estensioni sono di grande importanza, poiché sono spesso i veri obiettivi dello studio (Berger 2000). Ad esempio, in esperimenti clinici su nuovi farmaci l'obiettivo inferenziale è generalmente non ristretto a stabilire se il farmaco è attivo sui soggetti effettivamente osservati, ma a stabilire se è attiva nella sotto-popolazione indotta dalla procedura distorta, o nella popolazione obiettivo, o in entrambe, così da poter essere prescritta anche ad altri soggetti. Quanto a questo, Ludbrook and Dudley (1998) hanno un'opinione molto simile a quella espressa da Edgington: essi affermano che le estensioni inferenziali dai soggetti effettivamente osservati alla popolazione obiettivo dovrebbero essere « *verbal than statistical* », benché non danno indicazioni su come questo tipo di estensioni debbano essere fatte.

In Pesarin (2002) vengono discusse condizioni sufficienti che permettono di estendere, almeno debolmente, le inferenze condizionate a quelle incondizionate, anche quando i dati sono raccolti attraverso procedure distorte o quando, sotto l'ipotesi nulla, i modelli di verosimiglianza non posseggono *boundedly complete* o *invariant statistics* per tutte le entità di disturbo.

Infatti, quando le proprietà di similarità e di non distorsione condizionata sono congiuntamente soddisfatte e se correttamente applicabili, i test di permutazione



permettono tali estensioni in senso debole. E questo anche quando la sotto-popolazione indotta dalla vera procedura di selezione adottata e la popolazione obiettivo sono entrambe definite vagamente.

Con l'assunzione di scambiabilità sotto l'ipotesi nulla, le inferenze condizionate di permutazione hanno sempre una chiara interpretazione, mentre le loro estensioni deboli alla popolazione genitrice sottostante devono essere effettuate e interpretate con cautela. Queste estensioni e le interpretazioni a loro associate sono generalmente semplici e corrette quando i dati vengono raccolti con tecniche di campionamento casuale ben progettate da una data popolazione. Certamente, se i dati sono stati raccolti da procedure distorte, le estensioni non condizionate possono essere a volte ambigue e fuorvianti, sebbene siano generalmente appropriate e corrette per le popolazioni genitrici con rispetto alle quali i dati effettivi possono essere considerati come un campione casuale. In pratica, queste estensioni sono appropriate per le popolazioni selezionate indotte dalle specifiche procedure di campionamento distorte.

Si ritiene appropriato concludere questo capitolo con la seguente riflessione:

*«Io ho ammesso, come pure lo sperimentatore che deve sempre ammettere che è possibile di trarre valide inferenze dai risultati della sperimentazione; che è possibile risalire dalle conseguenze alle cause, dalle osservazioni alle ipotesi; come direbbe uno statistico dal campione alla popolazione da cui il campione è tratto, o come direbbe un logico dal particolare al generale. (...)*

*Noi dobbiamo subito riconoscere che ogni inferenza dal particolare al generale si raggiunge con un certo grado di incertezza, ma questo non è lo stesso di ammettere che tale inferenza non può essere assolutamente rigorosa perché la natura e il grado dell'incertezza sono per se stessi suscettibili di rigorosa espressione. Nella teoria delle probabilità, come è stata sviluppata nelle sue applicazioni ai giochi di sorte noi abbiamo il classico esempio per dimostrare questa possibilità. Se gli strumenti del giocatore sono realmente senza viziate le probabilità dei diversi possibili eventi o le combinazioni di eventi possono essere definiti con argomenti rigorosamente deduttivi, sebbene il presentarsi di ogni singolo risultato è riconosciuto come cosa incerta. Il solo fatto che le inferenze induttive sono incerte non può essere per ciò accettato come ostacolo per una inferenza perfettamente rigorosa e inequivocabile» (Fisher, 1935).*

# Capitolo 5

## I test di permutazione applicati agli studi sperimentali e osservazionali

### 5.1 INTRODUZIONE

Nel precedente capitolo è stato messo in evidenza che le assunzioni riguardanti la validità dei metodi parametrici vengono raramente soddisfatte nella pratica; pertanto, le inferenze che ne seguono, quando non improprie, sono necessariamente approssimate e le loro approssimazioni sono spesso difficili da valutare. Infatti, i metodi parametrici riflettono un approccio ad un modello e generalmente richiedono l'esplicita introduzione di specifiche restrizioni, che alle volte possono essere difficili da giustificare o anche abbastanza irrealistiche, se non su una base *ad hoc*, cosicché esse possono essere più specificatamente connesse con la disponibilità dei metodi che con la realtà. Al contrario, l'approccio non parametrico tenta di mantenere le assunzioni ad un più basso livello realizzabile, evitando quelle che sono difficili da giustificare o da interpretare e possibilmente senza un'eccessiva perdita di efficienza inferenziale.

Pertanto questo capitolo analizzerà i test di permutazione nella loro applicazione a due tipi di studi, quello sperimentale e quello osservazionale, per i quali, come si è discusso, risulta spesso difficile garantire le assunzioni previste dall'approccio parametrico.

### 5.2 STUDI SPERIMENTALI E OSSERVAZIONALI

Oscar Kempthorne (1979) scrive in riferimento agli studi sperimentali e osservazionali le seguenti affermazioni:

*«In un esperimento comparativo, perché si può congiungere solo un trattamento ad una qualunque unità sperimentale, non si può mai determinare la verità esatta e sicuramente si è confuso l'effetto del trattamento e l'effetto dell'unità sperimentale. Tuttavia, se si ha a disposizione un ampio esperimento e si è utilizzata la randomizzazione nel congiungere i trattamenti e le unità, si arriverà vicini alla verità che si potrebbe ottenere se si potesse congiungere ogni trattamento con ogni unità. Si otterrà una conclusione causale del tipo che imponendo un certo trattamento a un certo insieme di unità si otterrà una certa risposta».*

*«In uno studio osservazionale, al contrario, non si ha alcuna alternativa se non quella di esaminare, utilizzando l'analogia e l'analisi di dati, servendosi di un modello probabilistico. Il risultato può solo essere una descrizione dei dati attraverso un modello di probabilità, un modello che è conforme ai dati da come è giudicato da un test di goodness of fit. La realizzazione di inferenze causali è estremamente pericolosa».*

Da queste considerazioni emerge che gli studi osservazionali sono afflitti da una serie di problemi aggiuntivi rispetto a quelli sperimentali che possono pregiudicare la possibilità di condurre in tale ambito valutazioni di tipo inferenziale. Si ricordi che la logica degli studi osservazionali è quella di non interferire con la realtà osservata, ma di registrare ciò che si produce spontaneamente in essa. Nel seguito verranno discusse le argomentazioni teoriche caratterizzanti i due tipi di studi, quello sperimentale e quello osservazionale, mettendo in evidenza le difficoltà e le possibili soluzioni.

### **5.3 L'OBIETTIVO DEL CONFRONTO**

Quando l'interesse del ricercatore è orientato a valutare l'effetto di un dato fattore su un dato evento, la soluzione metodologica usualmente adottata consiste nel *confrontare* gruppi di pazienti caratterizzati da differenti livelli del fattore di cui si vuole valutare l'effetto, e nel *valutarne la differenza* rispetto a una o più caratteristiche o risposte che qualificano l'evento di interesse. Si considererà questa semplice pianificazione, cioè un confronto diretto di due gruppi che differiscono per un'ipotetica variabile causale al fine di paragonare gli studi sperimentali a quelli osservazionali.

### **5.4 LA COMPARABILITÀ TRA I GRUPPI**

Generalmente le valutazioni comparative non si limitano alla *descrizione* della differenza nei profili di risposta dei gruppi confrontati, ma procedono in senso più analitico nel verificare se è più verosimile che tale differenza sia dovuta semplicemente al caso e al fatto che è stato osservato solo un campione della popolazione di interesse o se al contrario è più verosimile che essa rifletta una reale diversità nei gruppi a prescindere dal campionamento, o in altri termini, che la differenza sia *statisticamente significativa*. Quest'ultimo punto rappresenta in molti casi l'obiettivo principale dello studio: interpretare la differenza delle risposte dei gruppi come effetto del trattamento, riconoscendo nell'associazione trattamento-risposta una relazione di tipo *causa-effetto*.

Una differenza nelle risposte dei gruppi confrontati può però prodursi non solo per effetto del trattamento. Vi possono essere infatti uno o più fattori, detti *fattori di confondimento*, di cui è noto il potenziale effetto sulle risposte dei soggetti e che se presenti con peso

diverso nei gruppi confrontati, possono contribuire a determinare le differenze osservate *confondendo* l'effetto del trattamento.

Un possibile modo per isolare l'effetto del trattamento e per poter attribuire le differenze riscontrate tra i gruppi solo al trattamento, è quello di rendere i gruppi confrontati *comparabili*, ovvero simili rispetto a tutte le condizioni che possono differenziare i gruppi, ad eccezione del fattore in studio, cioè il trattamento.

Qualsiasi confronto volto a valutare differenze nelle risposte attribuibili al trattamento, dovrà quindi prevedere il controllo accurato della *comparabilità dei gruppi*. Se non si tiene in adeguato conto la presenza di possibili fattori di confondimento, il confronto diretto tra i gruppi di trattamento può risultare *sistematicamente distorto* e portare ad attribuire erroneamente la possibile differenza tra i gruppi all'effetto del trattamento.

## 5.5 LA STRUTTURA DEGLI STUDI PER GLI EFFETTI CAUSALI

Si considererà il caso di due trattamenti, numerati 1 e 0, e di un campione di  $n$  unità sotto studio. L' $i$ -esima unità possiede una risposta  $r_{1i}$  se è assegnata al trattamento 1, e una risposta  $r_{0i}$  se è assegnata al trattamento 0. In questa formulazione, gli effetti del trattamento sono definiti come il confronto fra  $r_{1i}$  e  $r_{0i}$ , come ad esempio  $r_{1i} - r_{0i}$ . Ciascuna unità riceve solo un trattamento, così viene osservata  $r_{1i}$  oppure  $r_{0i}$ , ma non entrambe, così il confronto implica un certo grado di congetture. In un certo senso, la stima degli effetti causali dei trattamenti può essere interpretata come un problema di dati mancanti, dal momento che uno fra  $r_{1i}$  e  $r_{0i}$  viene perso. Questa formulazione è usata nella letteratura della programmazione degli esperimenti, per esempio nei libri di Fisher (1951) e Kempthorne (1952). L'assunto che vi sia un unico valore  $r_{it}$  corrispondente all'unità  $i$  ( $i = 1, \dots, n$ ) e al trattamento  $t$  ( $t = 0, 1$ ) è stato denominato assunto di *stable unit-treatment value* (Rubin, 1980a).

Per l' $i$ -esima unità delle  $n$  unità nello studio ( $i = 1, \dots, n$ ), sia  $z_i$  l'indicatore dell'assegnazione al trattamento, con  $z_i = 1$  se l' $i$ -esima unità viene assegnata al trattamento 1 e  $z_i = 0$  se viene assegnata al trattamento 0. Sia  $\mathbf{x}_i$  un vettore di misurazioni osservate prima del trattamento o covariate per l' $i$ -esima unità.

Il *propensity score*,  $e(\mathbf{x})$ , è la probabilità condizionata di assegnazione al trattamento 1 date le covariate osservate; cioè  $e(\mathbf{x}) = \Pr\{z = 1 \mid \mathbf{x}\}$  dove si assume

$$\Pr\{z_1, \dots, z_n \mid \mathbf{x}_1, \dots, \mathbf{x}_n\} = \prod_{i=1}^n e(\mathbf{x}_i)^{z_i} \{1 - e(\mathbf{x}_i)\}^{1-z_i}$$

Sebbene l'assunzione di indipendenza sia forte e non necessaria, semplifica la notazione e la discussione.

### 5.5.1 UN'ASSUNZIONE CRITICA: ASSEGNAZIONE AL TRATTAMENTO FORTEMENTE IGNORABILE<sup>1</sup>

Gli studi randomizzati e non randomizzati differiscono in due modi:

- (1) in uno studio randomizzato il *propensity score* è una funzione nota, mentre negli studi osservazionali il *propensity score* è una funzione quasi sempre non nota;
- (2) in uno studio randomizzato, con un'appropriata collezione dei dati,  $\mathbf{x}$  contiene tutte le covariate che verranno utilizzate dallo sperimentatore per determinare le probabilità di randomizzazione (Rubin, 1977); per esempio, in un'esperimento completamente randomizzato, le probabilità di randomizzazione sono costanti per tutte le unità così  $\mathbf{x}$  può essere il vettore nullo. In un esperimento randomizzato, inoltre, l'assegnazione al trattamento  $z$  e la risposta  $(r_1, r_0)$  sono note essere condizionatamente indipendenti dato  $\mathbf{x}$ ; cioè

$$\Pr\{r_1, r_0, z \mid \mathbf{x}\} = \Pr\{r_1, r_0 \mid \mathbf{x}\} \cdot \Pr\{z \mid \mathbf{x}\} \quad (5.1)$$

Questa condizione spesso non regge in un esperimento non randomizzato;

L'assegnazione al trattamento sarà detta essere *fortemente ignorabile* dato  $\mathbf{x}$  se regge (5.1) e  $0 < \Pr\{z = 1 \mid \mathbf{x}\} < 1 \quad \forall \mathbf{x}$ .

Dal momento che l'assegnazione al trattamento fortemente ignorabile ricopre un ruolo critico, è generalmente necessario in ogni studio osservazionale indagare se questo assunto è consistente con i dati ottenibili, e determinare la sensibilità delle conclusioni a violazioni plausibili della *strong ignorability*. Per esempio, in caso di risposte binarie  $(r_1, r_0)$ , Rosenbaum e Rubin (1983b) descrivono un metodo per valutare la sensibilità delle conclusioni per certe deviazioni dalla *strong ignorability*; in particolare, l'assegnazione al trattamento viene assunta fortemente ignorabile dati  $\mathbf{x}$  e una covariata non osservata  $u$ , ma non fortemente ignorabile dato solo  $\mathbf{x}$ .

### 5.5.2 LA COMPARABILITÀ TRA GRUPPI NEGLI STUDI RANDOMIZZATI

Il problema della comparabilità dei gruppi viene risolto in maniera molto efficiente nell'ambito degli studi randomizzati. Questo aspetto contribuisce difatti a rendere la

---

<sup>1</sup> *strongly ignorable treatment assignment*, Rosenbaum (1984);

sperimentazione controllata, il contesto ideale per il confronto tra gruppi di trattamento finalizzato alla valutazione dell'efficacia di un trattamento. La sperimentazione controllata affronta direttamente o indirettamente il problema della comparabilità prima della fase di analisi dei dati.

Per esemplificare ciò, si espongono brevemente alcune considerazioni presentate da Fisher (1935) nel discutere il libro *“Gli effetti della fecondazione incrociata e dell'auto-fecondazione nel regno vegetale”* di Charles Darwin, in cui lo scopo dell'esperimento è di determinare se le differenze di origine tra piante auto-fecondate e piante incrociate influenzano il grado di sviluppo misurato dall'altezza a una certa data.

Fisher presenta il criterio dell'appaiamento finalizzato ad ottenere la confrontabilità dei gruppi rispetto ai fattori di cui siano note le influenze sulle risposte nel seguente modo: *«...si è visto che il metodo di comparazione adottato da Darwin è quello di contrapporre una pianta autofecondata a una etero-fecondata (incrociata), in condizioni rese il più possibili eguali. Le coppie così scelte per la comparazione erano germinate nello stesso tempo e le condizioni del suolo in cui esse si svilupparono erano largamente eguagliate col piantarle nello stesso vaso. (...)»<sup>2</sup> L'evidente obiettivo di queste precauzioni è di aumentare la sensibilità dell'esperimento, rendendo le differenze nel grado di sviluppo che devono essere osservate il meno possibile dipendenti dalle circostanze ambientali e il più possibile, perciò, dalle intrinseche differenze dovute al loro modo di origine. (...) La procedura dell'appaiamento, come in verità era il suo proposito, ha eguagliato alcune differenze nelle condizioni del suolo, illuminazione, correnti d'aria, ecc. in cui parecchie coppie di individui possono differire».*

A questo punto Fisher esamina le condizioni fisiche della tecnica sperimentale richieste per garantire l'assunzione che, se la differenza di origine è ininfluente, il risultato dell'esperimento sarà completamente governato dalla legge del caso. L'elemento che nel procedimento sperimentale costituisce una essenziale salvaguardia è costituito dalla randomizzazione, poiché le cause incontrollate che possono influenzare il risultato, sono, a rigore, sempre innumerevoli: *«...Se ora, quando le quindici coppie<sup>3</sup> di posti sono stati scelti e così facendo tutte le differenti paia saranno esposte durante il corso dell'esperimento, sono state predeterminate, noi assegnamo a caso col lancio di una moneta, quale posto deve essere occupato dalle piante incrociate e quale da quelle autofecondate, noi con lo stesso atto determineremo se questo particolare ingrediente dell'errore apparirà nella nostra media con segno positivo o negativo.*

*Poiché ogni particolare errore ha così una eguale e indipendente probabilità di essere positivo o negativo l'errore della nostra media sarà necessariamente distribuito secondo una distribuzione di frequenza, centrata a zero che sarà simmetrica nel senso che a ogni*

<sup>2</sup> Fisher inoltre aggiunge fra le righe *«...i genitori presumibilmente provenivano dalla stessa partita di semi»;*

<sup>3</sup> la serie di piante considerata nell'esperimento storico sul grado di sviluppo è la *Zea Mais*; di essa, sono state osservate 15 coppie di piante, ogni coppia comprendente una pianta autofecondata e una incrociata;

*possibile errore positivo vi corrisponde un eguale errore negativo, e questi, come la nostra procedura garantisce, si presenteranno in effetti con eguale probabilità.(...)*

*Una randomizzazione fatta in modo appropriato, cioè in modo che ogni coppia di piante abbia assegnata la sua posizione indipendentemente a caso, assicura che la stima dell'errore tenga debito conto di tutte le cause del differente grado di sviluppo e alleggerisce lo sperimentatore dalla preoccupazione di considerare e stimare la grandezza delle innumerevoli cause per cui i suoi dati possono essere disturbati. L'unico difetto nella procedura di Darwin era l'assenza della randomizzazione».*

In Arboretti *et al.* (2000), con riferimento agli studi clinici randomizzati (RCT), vengono descritti i seguenti criteri al fine di affrontare il problema della comparabilità (tabella 5.1):

---

**Tabella 5.1:** criteri per rendere i gruppi comparabili nell'ambito degli studi clinici randomizzati (Arboretti *et al.*, 2000)

---

1. *criteri di inclusione dei soggetti nello studio:* pensati per massimizzare la possibilità di far emergere l'efficacia del trattamento considerando i pazienti che hanno un'appropriata indicazione clinica per la terapia, contribuiscono indirettamente a rendere meno eterogenei i soggetti inclusi nello studio;
  2. *stratificazione/appaiamento:* stratificando o appaiando a priori i pazienti rispetto ad alcuni fattori (es. età, sesso) di cui siano note le influenze sulla patologia oggetto di studio, si ottiene la confrontabilità dei gruppi rispetto a tali fattori, all'interno degli strati o degli insiemi appaiati costruiti;
  3. *studio cieco o doppio cieco:* la non conoscenza da parte del paziente e/o del medico del trattamento assegnato, evita l'introduzione di distorsioni conscie ed inconscie nel processo di assegnazione degli individui ai gruppi di trattamento, impedendo così in particolare che i soggetti con particolari caratteristiche cliniche o prognostiche, che possono influire sulle risposte, siano assegnati a determinati trattamenti;
  4. *randomizzazione:* l'assegnazione casuale dei soggetti ai diversi livelli di trattamento tende a rendere comparabili i gruppi di trattamento rispetto a tutti i fattori, noti e ignoti, misurabili e non, che possono influire sulle risposte dei soggetti, eccetto che per quello in studio;
  5. *monitoraggio costante dell'andamento dello studio:* la verifica costante dello svolgimento dello studio può limitare problemi riguardanti l'applicazione corretta delle modalità previste per l'acquisizione dei dati e la qualità di quest'ultimi;
- 

### 5.5.3 LA COMPARABILITÀ TRA GRUPPI NEGLI STUDI OSSERVAZIONALI

Negli studi osservazionali, con riferimento per esempio ad un ambito ospedaliero, il problema della comparabilità presenta una complessità maggiore rispetto al contesto della sperimentazione clinica, complessità che spiega la problematicità dell'applicazione delle procedure inferenziali per il confronto dei diversi gruppi di trattamento e della conseguente interpretazione *causale* dei risultati.

In particolare, con riferimento agli aspetti del problema della comparabilità legati al trattamento, negli studi osservazionali quest'ultimo non è assegnato ai pazienti sotto il controllo del ricercatore e mediante un meccanismo di randomizzazione, ma in una varietà di condizioni incontrollate. Può essere scelto dal soggetto stesso, o dal medico, può essere



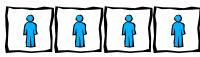
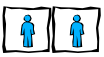


influenzato da un certo numero di fattori, inclusi l'indicazione per il trattamento, la prognosi del paziente, altre caratteristiche del paziente (es. età, sesso), le preferenze del paziente. Oltre ai fattori che si riferiscono alle caratteristiche personali dei soggetti, ve ne sono altri che riguardano le modalità di gestione clinica-terapeutica-assistenziale da parte del personale curante e altri ancora relativi alle modalità di registrazione e di raccolta delle informazioni. In uno studio osservazionale quindi, ci si aspetta che i gruppi posti a confronto differiscano rispetto ad un elevato numero di fattori oltre a quello di cui si vuole valutare l'effetto, cioè il trattamento.

## 5.6 IL CONTROLLO DEI FATTORI DI CONFONDIMENTO IN FASE DI ANALISI

Negli studi osservazionali è rilevante l'applicazione successiva alla raccolta dei dati di metodi statistici per il controllo dei fattori di confondimento *noti e rilevati* nel corso dello studio. A tale riguardo si può osservare che tra i fattori che possono confondere l'effetto del trattamento, ve ne sono alcuni rappresentanti caratteristiche facilmente osservabili e quantificabili del campione in studio; per queste variabili si può prevedere l'applicazione di metodi statistici con il fine di aggiustare *a posteriori* per differenze tra i gruppi rispetto a tali fattori. Le tecniche statistiche più comunemente usate per questo scopo sono rappresentate da “metodi di stratificazione” e “metodi basati su modelli prognostici”.

I metodi di *stratificazione a posteriori* prevedono che i soggetti osservati siano raggruppati in *strati* sulla base di uno o più fattori di confondimento. L'idea alla base di tali metodi è che eventuali differenze tra i gruppi osservati, possano essere rimosse confrontando i soggetti sottoposti ai vari livelli del trattamento, che presentano gli stessi valori di tali fattori, cioè soggetti appartenenti allo stesso strato (es. tabella 5.2). I risultati ottenuti per ciascuno strato vengono poi riassunti in un'unica misura relativa all'intero campione. Il limite principale di tali metodi riguarda le difficoltà pratiche che insorgono per il controllo di molti fattori di confondimento.

**Tabella 5.2:** il gruppo esposto al trattamento e quello non esposto sono molto sbilanciati rispetto al sesso, variabile che si suppone possa influenzare le risposte dei soggetti. I pazienti vengono quindi stratificati in due strati rispetto al sesso, e il confronto prosegue all'interno di ciascun strato. Se si evidenziano differenze nelle risposte dei soggetti, si è sicuri che queste non sono imputabili al sesso. (Arboretti *et al.*, 2000)

	<i>esposti</i>	<i>non esposti</i>
<i>campione totale</i>		
<i>strato 1</i>		
<i>strato 2</i>		



I metodi basati su *modelli prognostici* prevedono che il controllo dei fattori di confondimento sia effettuato mediante l'introduzione di un modello statistico (es. lineare, logistico) in cui le risposte dei soggetti sono espresse in funzione del trattamento e dei fattori di confondimento. Il modello ingloba quindi il fattore di trattamento e tutte le variabili di confondimento note che hanno un valore prognostico (tabella 5.3).

<i>risposta</i>	=	<i>fattore di interesse</i>	+	<i>fattori di confondimento</i>	...	<i>fattori di confondimento</i>
○		○		○		○

**Tabella 5.3:**  
modello prognostico (Arboretti *et al.*, 2000)

Riguardo ai vantaggi presentati, tali metodi consentono il controllo di un numero elevato di fattori di confondimento. Con riferimento ai limiti, se il modello proposto (es. lineare) non è appropriato a rappresentare il tipo di legame esistente tra la risposta e i fattori di confondimento, nonché le relazioni tra il trattamento e le variabili di confondimento incluse nel modello, non si ha un adeguato controllo del confondimento.

Un ulteriore procedura che si rivela molto efficiente per il controllo dei fattori di confondimento è rappresentata dalla tecnica del *propensity score*, che verrà presentata nel paragrafo successivo.

#### 5.6.1 UN'APPLICAZIONE NEGLI STUDI OSSERVAZIONALI CON STRATIFICAZIONE A POSTERIORI PER *PROPENSITY SCORE*

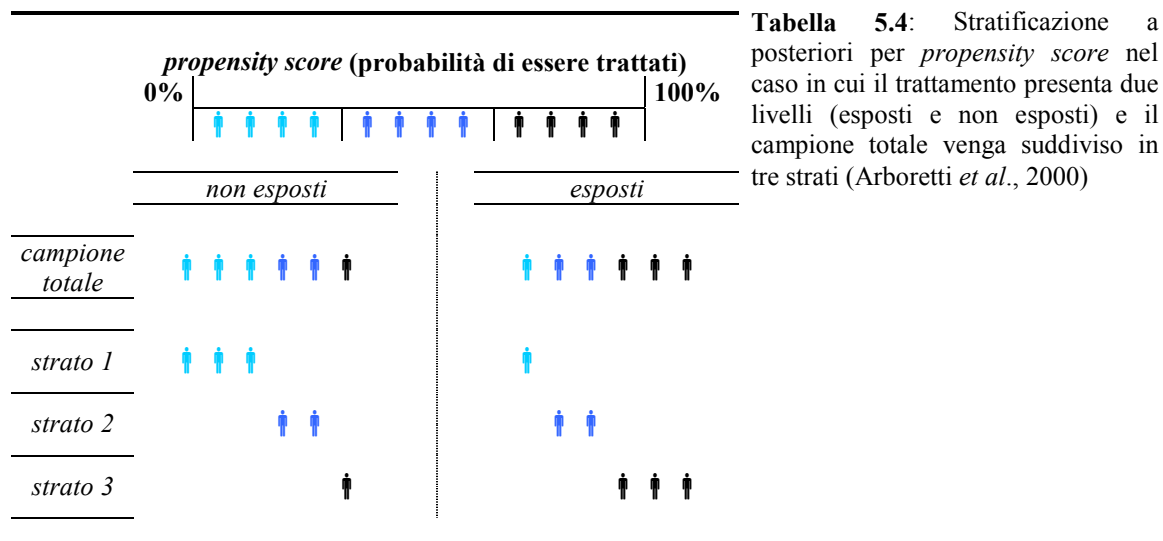
L'idea alla base della tecnica del *propensity score* è di sostituire all'insieme dei fattori di confondimento, una loro *sintesi* che viene poi utilizzata come singola variabile di confondimento. Tale metodo quindi, riassume l'insieme delle variabili di confondimento in un unico fattore, consentendo di ricondursi ad un problema di confronto tra due o più gruppi di trattamento in presenza di un unico fattore di confondimento.

Si considerino per semplicità espositiva, due livelli di trattamento che indichino rispettivamente l'esposizione e la non esposizione al trattamento. Il *propensity score* può essere stimato dai dati osservati, utilizzando ad esempio un modello (es. modello logistico) in cui l'esposizione al trattamento è la variabile dipendente e i fattori di confondimento rappresentano le variabili esplicative. Per ciascun paziente viene quindi calcolato un valore compreso fra 0 e 1 che esprime la probabilità di essere assegnati al gruppo di trattamento, ovvero la propensione ad appartenere al gruppo dei trattati, determinata come funzione dei fattori di confondimento considerati.

Una volta sintetizzata la quantità di confondimento in un singolo *score*, si può prevedere ad esempio<sup>4</sup> una stratificazione a posteriori, formando degli strati definiti in

<sup>4</sup> la stratificazione a posteriori per *propensity score* è presentata in Arboretti *et al.*, (2000);

corrispondenza di intervalli di valore del *propensity score*. Se il modello che stima la propensione al trattamento, cattura interamente la relazione tra l'insieme dei fattori di confondimento e il fattore di trattamento, allora all'interno di ciascun strato, che include pazienti con valori simili del *propensity score* e quindi con probabilità di assegnazione al trattamento all'incirca uguale, soggetti trattati e di controllo tendono ad essere simili rispetto a tutti i fattori di confondimento inclusi nella stima del *propensity score* (tabella 5.4):



**Tabella 5.4:** Stratificazione a posteriori per *propensity score* nel caso in cui il trattamento presenta due livelli (esposti e non esposti) e il campione totale venga suddiviso in tre strati (Arboretti *et al.*, 2000)

La stratificazione per *propensity score* combina gli aspetti positivi della stratificazione e del modello prognostico, consentendo il controllo di un numero elevato di fattori di confondimento. In particolare, la stratificazione per *propensity score* consente il controllo di un numero elevato di fattori di confondimento.

Infatti, una tecnica standard per l'aggiustamento negli studi osservazionali viene presentata in Cochran (1963) sotto il nome di *subclassification*. Questo metodo permette di aggiustare per le variabili di confondimento dividendo sulla base di  $x$  le unità sperimentali e di controllo in sotto-classi o strati. Come metodo di aggiustamento multivariato, la subclassificazione ha il vantaggio di implicare confronti diretti di gruppi apparentemente comparabili all'interno di ogni sotto-classe; un problema che ne consegue è che all'aumentare delle variabili di confondimento, il numero delle sotto-classi cresce drammaticamente, cosicché anche con solo due categorie per variabile, si otterranno  $2^P$  sotto-classi per  $P$  variabili, con molte delle sotto-classi che non conterranno né le unità di controllo né le unità di trattamento. La subclassificazione per *propensity score* è un modo naturale per evitare questa difficoltà.

Tra gli aspetti importanti da considerare per un'applicazione corretta di tale metodo vi è la scelta dei fattori di confondimento da utilizzare per la stima del *propensity score*. Tale scelta richiede una elevata padronanza dell'argomento di studio per giudicare quali potenziali fonti di confondimento sono più rilevanti e devono essere in qualche modo

controllate, quali sono quelle meno rilevanti e quindi eventualmente da ignorare, poiché al crescere dei fattori di confondimento aumenta la complessità dell'analisi. Non esiste un criterio unico al riguardo. Generalmente l'approccio che viene seguito è quello di considerare inizialmente una lista di fattori sulla base delle conoscenze scientifiche circa il loro ruolo come fonti di confondimento, insieme a confronti esplorativi per verificare quali fattori presentano un peso diverso nei vari gruppi di trattamento. A questo punto si stima il *propensity score* e si formano gli strati. All'interno degli strati così formati si controlla se i gruppi da confrontare risultano o meno omogenei. Se sussistono ancora degli sbilanciamenti, il modello per la stima del *propensity score* può essere di conseguenza riformulato.

Un'ultima considerazione importante da sottolineare è che il bilanciamento dei gruppi di trattamento rispetto a potenziali fattori di confondimento ottenuto utilizzando i metodi basati sul *propensity score*, differisce dal bilanciamento che deriva dall'applicazione del meccanismo della randomizzazione negli studi sperimentali. La randomizzazione tende infatti a bilanciare, ovvero a rendere simili i gruppi di trattamento, rispetto a *tutte* le variabili di confondimento, osservate o non osservate. Negli studi osservazionali, i metodi basati sul *propensity score* hanno la proprietà di bilanciare i gruppi solo rispetto alle variabili di confondimento osservate e utilizzate per stimare il *propensity score*. L'applicazione di tali metodi deve essere di conseguenza accompagnata da un'attenta valutazione della possibile presenza di ulteriori fattori di confondimento non adeguatamente considerati.

## 5.7 DALL'ASSOCIAZIONE TRATTAMENTO-RISPOSTA ALLA RELAZIONE DI TIPO CAUSA-EFFETTO

Considerato il problema dei fattori di confondimento, l'analisi procede con il confronto tra i gruppi di trattamento per verificare se la differenza tra i profili di risposta è da considerarsi dovuta al caso o se è al contrario *statisticamente significativa* e se è possibile attribuirle, in quest'ultimo caso, all'effetto del trattamento. Il procedimento statistico che consente di decidere sulla base dei dati del nostro campione è la *verifica di ipotesi*.

Si parla generalmente di *inferenza causale* quando nella valutazione della significatività statistica di una differenza tra i gruppi, entra in gioco la possibile *spiegazione* di tale differenza.

Nella sperimentazione controllata, il ricercatore che vuole conoscere gli effetti di un trattamento, lo può applicare, e con la semplice precauzione della randomizzazione, avrà garanzia della validità della prova di significatività e pertanto può ottenere una risposta diretta alla sua domanda.

La possibilità di interpretare in senso causale relazioni tra il trattamento e la risposta nell'ambito degli studi osservazionali dipende molto dal controllo adeguato dei fattori di

confondimento. Ma anche nel caso in cui il ricercatore abbia posto la massima attenzione in questa fase, non si può escludere che la differenza riscontrata tra i gruppi, non sia dovuta totalmente o in parte all'effetto del trattamento, ma a qualche fattore ignoto o non rilevato di cui non si è tenuto conto nel controllo dei fattori di confondimento.

Come procedere alla luce di queste considerazioni? Un importante contributo è dato da alcune considerazioni espresse da Cochran (1963) in un lavoro sulla pianificazione e analisi degli studi osservazionali. L'autore cita la semplice e acuta osservazione di Sir Ronald Fisher che diversi anni prima, alla domanda. *«Come operare il passo dall'associazione alla causalità per dati di natura osservazionale?»* rispondeva: *«rendete elaborate le vostre teorie scientifiche»*. Nella risposta di Fisher confluiscono considerazioni e implicazioni che attengono al disegno dello studio e al momento inferenziale, così commentate da Cochran: *«...Nel costruire un'ipotesi causale sull'effetto del trattamento, il ricercatore dovrebbe immaginare quante più possibili conseguenze osservabili se l'ipotesi fosse vera e pianificare di conseguenza studi osservazionali con l'obiettivo di valutare se ciascuna di queste conseguenze si verifichi o meno»*.

L'ipotesi di ricerca diventa in quest'ottica, “elaborata” ovvero estremamente particolareggiata, sia al momento della sua definizione che al momento della verifica statistica con i dati osservati. Il ricercatore anticipa in modo accurato tutte le conseguenze attese nell'ipotesi che il trattamento abbia effetto, specificando in che modo e su quali aspetti si ipotizza che il trattamento abbia effetto. Il risultato è quindi la definizione dettagliata di un dettagliato pattern *multivariato* (in quanto si valuta l'effetto su più aspetti o più “variabili risposta”) di risultati attesi *coerente* con l'ipotesi di effetto causale del trattamento.

Perché risulta importante la definizione di un dettagliato pattern coerente di risultati attesi nell'ambito degli studi osservazionali? Se tale pattern viene confermato nei dati, l'evidenza a favore della spiegazione causale dell'associazione trattamento-risposta è forte perché ogni possibile spiegazione alternativa a questa ipotesi e quindi ogni affermazione che il trattamento non ha effetto, deve poter proporre una spiegazione alternativa per l'intero pattern. La definizione di un pattern dettagliato coerente non elimina certo la possibilità di spiegazioni alternative a quella avanzata dai ricercatori, ma ne riduce sicuramente il numero, rendendo quindi al contempo più forte l'evidenza dei risultati a favore di una spiegazione causale.

In concetto di “teoria elaborata” discusso da Cochran e Fisher, non riguarda solo il momento della definizione dell'ipotesi di ricerca e della sua verifica statistica dei dati. Tale concetto ha importanza anche nella fase di interpretazione dei risultati derivanti dall'applicazione del test di significatività. Gli autori difatti aggiungono: *«...Nell'interpretare i risultati di uno studio osservazionale che mostra un'associazione consistente con l'ipotesi causale, il ricercatore dovrebbe ricercare e discutere tutte le possibili spiegazioni alternative dei risultati ottenuti, incluse ipotesi differenti da quella ipotizzata e/o possibili distorsioni verificatesi nel corso dello studio»*. La specificazione di

un pattern multivariato di risultati coerente con l'ipotesi di effetto causale del trattamento può congiuntamente ad un'attenta discussione dei risultati, ridurre il numero di tali spiegazioni alternative. Ma la natura osservazionale dello studio pone comunque un veto alla possibilità di conclusioni *confirmative* sull'effetto del trattamento.

L'evidenza derivata da uno studio osservazionale ben condotto è sì più debole, più problematica rispetto a quella risultante da uno studio sperimentale, ma può contribuire in maniera importante, congiuntamente ad altre, allo studio di relazioni di tipo causa-effetto.

## 5.8 I TEST DI IPOTESI

Lo strumento statistico che consente di operare una verifica di ipotesi, ovvero di decidere sulla base dei dati del nostro campione se l'ipotesi di ricerca alla base del confronto tra i gruppi è o meno accettabile, è detto *test di ipotesi*. Mediante il test di ipotesi viene calcolato il livello di significatività statistica noto con il termine di *p-value*. Se tale valore risulta inferiore al livello di significatività prefissato dal ricercatore (che rappresenta il margine di errore accettabile insito nel procedimento), allora si rifiuta l'ipotesi che la differenza osservata sia casuale e si conclude affermando che la differenza è statisticamente significativa al livello di significatività prefissato.

### 5.8.1 I TEST DI PERMUTAZIONE APPLICATI AGLI STUDI SPERIMENTALI

Ancora con riferimento all'esperimento storico sul grado di sviluppo descritto in precedenza, si rifletta attraverso le osservazioni di Fisher su quale sia l'ipotesi nulla da verificare: «...*Lo scopo dell'esperimento è di determinare se le differenze di origine tra piante auto-fecondate e incrociate influenzano il loro grado di sviluppo misurato dall'altezza a una certa data. In altre parole se il numero delle piante fosse aumentato indefinitamente, il nostro scopo è di determinare se le altezze medie a cui questi due aggregati di piante tendono sono eguali o ineguali. La più generale affermazione della nostra ipotesi zero, è, perciò, che i limiti a cui queste due medie tendono sono uguali*». Fisher descrive come il test *t* di Student, in conformità con la classica teoria degli errori, risulti appropriato all'ipotesi nulla che i due gruppi di misure sono campioni ricavati dalla stessa popolazione normalmente distribuita: «...*La teoria degli errori ci mette in grado di provare una ipotesi alquanto più limitata che, per larga esperienza è stata trovata appropriata ai caratteri metrici del materiale sperimentale in biologia. Le cause di disturbo che introducono discrepanze delle medie di misurazioni di materiale simile, sono state trovate tali da produrre effetti che si conformano in modo soddisfacente alla distribuzione teoretica nota come legge normale di frequenza degli errori. E' questa circostanza che rende appropriato lo scegliere, come ipotesi zero da provare, una per la quale un esatto criterio statistico è utilizzabile, cioè quella che i due gruppi di*

*misurazione sono campioni ricavati dalla stessa popolazione normale. Sulla base di questa ipotesi noi possiamo procedere a confrontare le differenze medie di altezza tra piante incrociate e autofecondate con le differenze che potrebbero essere aspettate tra queste medie, tenendo conto delle discrepanze osservate tra le altezze delle piante della stessa origine».*

In base alle considerazioni iniziali sull'ipotesi più generale, Fisher propone di estendere il requisito della normalità nel seguente modo: «...Potrebbe nondimeno essere legittimo domandarsi se si dovessero ottenere i risultati materialmente differenti qualora fosse possibile provare la più larga ipotesi la quale semplicemente asserisce che le due serie sono ricavate dalla stessa popolazione, senza specificare che questa è normalmente distribuita. (...) L'azione fisica della randomizzazione (...) procura la possibilità, rispetto a ogni particolare complesso di dati, di esaminare la più larga ipotesi in cui non è implicita la normalità della distribuzione. (...) In base all'ipotesi che le due serie di semi sono campioni presi a caso da identiche popolazioni e che i loro posti sono stati assegnati ai membri di ogni paio indipendentemente a caso, le 15 differenze sarebbero capitate con egual frequenza di segno positivo o di segno negativo. (...) Poiché "ex hypothesi" ognuna di queste  $2^{15}$  combinazioni capiterà per caso con egual frequenza, una conoscenza di quante di esse sono uguali e più grandi del valore effettivamente osservato, offre una diretta prova aritmetica della significatività di questo valore».

La conclusione che ne consegue è la seguente: «*The use of permutation inferences in randomized experiments is sustained or warranted by the physical act of randomization, which creates the only probability distribution used in the inference*» (Fisher, 1935). Nel seguito si vuole formalizzare questa conclusione con riferimento alla struttura degli studi per gli effetti causali presentata nel paragrafo 5.5.

*Osservazione 5.1:* Si noti che un errore di II tipo (accettare l'ipotesi nulla quando essa è falsa) può avere serie conseguenze alla pari di un errore di I tipo (rifiutare la vera ipotesi nulla) se un errore di II tipo viene considerato effettivamente un errore, e cioè, se il fallire il rifiuto dell'ipotesi nulla viene interpretato come lo stabilire la veridicità dell'ipotesi nulla. Si deve tener ben presente che l'ipotesi nulla non sarà mai provata o accertata (non potrà mai essere provata da un numero finito di esperimenti). «*Ogni esperimento si può dire che ha lo scopo di dare ai fatti una occasione di confutare l'ipotesi nulla*» (Fisher, 1935), ma non sarà mai una confutazione certa. Ad ogni modo, un errore di I tipo viene raramente riconosciuto come tale, e vi è la possibilità che i risultati vengano accettati dalle apparenze (rendendo improbabile che lo studio sia ripetuto). Nella ricerca biomedica, per di più, si ha una preferenza maggiore nel controllare l'errore di I tipo, poiché la punizione per un'inferenza falsamente positiva può essere l'introduzione di una nuova terapia senza valore. Come tale, una proprietà chiave della verifica d'ipotesi è l'esattezza (Berger, 2000), o almeno la conservatività; al contrario, per quanto visto nel paragrafo 4.3 e per

quanto si vedrà nel paragrafo 7.9, il venir meno ad alcune assunzioni può rendere un test anticonservativo.

### 5.8.2 IL TEST DI PERMUTAZIONE DI FISHER NEGLI ESPERIMENTI RANDOMIZZATI

Il test di permutazione di Fisher (1935) esamina l'ipotesi nulla puntuale di nessuna differenza negli effetti dei trattamenti per ogni unità sperimentale; cioè

$$H_0 : r_{1i} = r_{0i} \quad \text{per } i = 1, \dots, n \quad (5.2)$$

Si noti che questa ipotesi nulla puntuale afferma che la stessa risposta dovrebbe venir osservata per ciascuna unità che ha ricevuto il trattamento alternativo.

Sia  $r_i^*$  la risposta osservata dell' $i$ -esima unità, cioè  $r_i^* = z_i \cdot r_{1i} + (1 - z_i) \cdot r_{0i}$ , e si denoti il vettore delle risposte osservate con  $\mathbf{r} = (r_1^*, r_2^*, \dots, r_n^*)^T$  e il vettore delle assegnazioni al trattamento con  $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$ . Sia  $t(\mathbf{z}, \mathbf{r})$  una statistica scelta per misurare le deviazioni dall'ipotesi nulla puntuale; per esempio,  $t(\mathbf{z}, \mathbf{r})$  potrebbe essere la differenza in media campionaria delle risposte dei due trattamenti, cioè

$$t(\mathbf{z}, \mathbf{r}) = \{\mathbf{z}^T \mathbf{r} / \mathbf{z}^T \mathbf{1}\} - \{(\mathbf{1} - \mathbf{z})^T \mathbf{r} / (\mathbf{1} - \mathbf{z})^T \mathbf{1}\}$$

dove  $\mathbf{1}$  è un vettore  $n$ -dimensionale di 1.

Fisher propose la verifica della ipotesi nulla puntuale usando le code della permutazione di  $t(\mathbf{z}, \mathbf{r})$  indotte dalla randomizzazione, dove  $\mathbf{r}$  è, in un certo senso, trattata come una costante.

Si vuole ora riesaminare la giustificazione per il condizionamento alla risposta osservata  $\mathbf{r}$  nel test di permutazione di Fisher. In un esperimento completamente randomizzato, l'assegnazione al trattamento  $z$  ha una distribuzione nota, cioè indipendente dalla risposta  $(r_1, r_0)$  e inoltre  $0 < \Pr\{z = 1\} < 1$ ; l'assegnazione al trattamento è quindi fortemente ignorabile senza alcuna covariata, cioè, con  $\mathbf{x}$  pari al vettore nullo. La distribuzione della risposta osservata,  $r^* = r_z$ , generalmente dipende da  $z$ ; comunque sotto l'ipotesi nulla puntuale, la risposta osservata soddisfa a  $r_z = r_1 = r_0$ , così l'assegnazione al trattamento  $z$  è indipendente dalla risposta osservata  $r_z$ . Sotto l'ipotesi nulla, quindi, la distribuzione condizionata delle assegnazioni al trattamento date le risposte osservate,  $\Pr\{\mathbf{z} | \mathbf{r}\}$ , è uguale alla distribuzione marginale di randomizzazione delle assegnazioni al trattamento,  $\Pr\{\mathbf{z}\}$ . Pertanto, sotto l'ipotesi nulla, la distribuzione condizionata della statistica test  $t(\mathbf{z}, \mathbf{r})$  dato il valore delle risposte osservate,  $\mathbf{r} = \mathbf{c}$ , eguaglia la distribuzione di permutazione di  $t(\mathbf{z}, \mathbf{c})$  indotta dalla randomizzazione. Formalmente, per ogni costante  $\mathbf{c}$ ,

$$\Pr\{t(\mathbf{z}, \mathbf{r}) \mid \mathbf{r} = \mathbf{c}\} = \Pr\{t(\mathbf{z}, \mathbf{c}) \mid \mathbf{r} = \mathbf{c}\} = \Pr\{t(\mathbf{z}, \mathbf{c})\} \quad (5.3)$$

dalla (5.2) e (5.1) con  $\mathbf{x}$  pari al vettore nullo. La conclusione è che il test di Fisher ha la *size* condizionata corretta dato  $\mathbf{r}$  osservato, e pertanto anche la *size* non condizionata corretta, come immediata conseguenza della (5.3).

Si noti che la giustificazione per il test di Fisher si appoggia a due condizioni:

- (i) la distribuzione di permutazione di  $\mathbf{z}$ , e quindi anche di  $t(\mathbf{z}, \mathbf{c})$  per ogni costante  $\mathbf{c}$ , è nota, dal momento che viene creata dallo sperimentatore;
- (ii) l'assegnazione al trattamento è fortemente ignorabile senza covariate, così sotto l'ipotesi nulla puntuale, la distribuzione nota di permutazione di  $t(\mathbf{z}, \mathbf{c})$  è eguale alla pertinente distribuzione condizionata di  $t(\mathbf{z}, \mathbf{r})$  date le risposte osservate  $\mathbf{r} = \mathbf{c}$ ; cioè regge (5.3). Negli studi osservazionali, anche se l'assegnazione al trattamento è fortemente ignorabile, la distribuzione delle assegnazioni al trattamento è generalmente non nota, e pertanto il test di randomizzazione di Fisher non è generalmente applicabile.

### 5.8.3 I TEST DI PERMUTAZIONE APPLICATI AGLI STUDI OSSERVAZIONALI

Come precedentemente discusso nel paragrafo 4.3, molti test usuali necessitano che siano soddisfatte alcune assunzioni sulle caratteristiche della popolazione da cui il campione è tratto e inoltre chiedono che quest'ultimo costituisca un *campione casuale* dalla popolazione di interesse, ma come si è visto questa condizione viene raramente soddisfatta.

Infatti negli studi sperimentali è comune che le unità da trattare non siano state scelte casualmente dalla popolazione, bensì siano state selezionate in qualche modo fra quelle disponibili e *randomly assigned* ai livelli stabiliti del trattamento. Negli studi osservazionali vi è una propensione minore a introdurre assunzioni e ipotesi sulla popolazione di provenienza del campione, inoltre il campionamento casuale da una data popolazione è difficilmente e raramente realizzabile nelle applicazioni reali. La natura osservazionale dello studio pone inoltre ulteriori problemi: possibile presenza di fattori di confondimento, possibilità di numerosità ridotte dei gruppi confrontati, possibile presenza di valori mancanti, ecc.

Con riferimento alla discussione della “teoria elaborata” presentata nel paragrafo 5.7, l'attenzione è rivolta all'individuazione di un test per verificare l'ipotesi nulla di assenza di relazione trattamento-risposta contro l'ipotesi alternativa che il trattamento abbia effetto sulla risposta secondo il pattern coerente specificato dal ricercatore e che per questo verrà definita *ipotesi alternativa coerente*.



L'operazione non è però così semplice, in quanto il pattern coerente può presentare una complessità tale da rendere estremamente arduo il confronto tra i gruppi. Il trattamento può avere effetto su diversi tipi di variabili qualitative o quantitative; inoltre per ciascuna variabile risposta viene specificata la *direzionalità* dell'effetto (riduzione o aumento). Ci si può trovare quindi ad affrontare problemi complessi di verifica di ipotesi "multivariata" (più variabili risposta) con ipotesi alternative "direzionali" (direzionalità dell'effetto).

Una soluzione ai vari problemi che insorgono nel contesto osservazionale, è data dall'applicazione di test per la verifica di ipotesi basati sulla *combinazione non parametrica di test di permutazione dipendenti*<sup>5</sup> (NPC)(Pesarin, 2001), una metodologia questa particolarmente efficiente che consente di risolvere problemi complessi di verifica di ipotesi, ad esempio in presenza di molte variabili risposta, e/o dati mancanti, e/o nei casi di campionamento non casuale.

Ad esempio, con riferimento al paragrafo 5.6.1, la NPC consente di effettuare un test per il confronto tra gruppi stratificati rispetto a valori del *propensity score* mediante la scomposizione del problema di verifica in tre fasi:

- 1) all'interno di ciascun strato, vengono simultaneamente applicati più test detti *test parziali*, uno per ogni variabile risposta considerata. Ciascun test parziale è volto ad esaminare il contributo di ogni singola variabile risposta considerata;
- 2) All'interno di ciascun strato, i test parziali vengono combinati in un unico test detto *test combinato di secondo ordine*, per verificare separatamente per ciascun strato, se vi siano delle differenze nei gruppi rispetto all'insieme delle variabili risposta considerate globalmente;
- 3) Si combinano i test del secondo ordine relativi ai vari strati in un unico *test combinato del terzo ordine*, relativo all'ipotesi globale di differenza tra i gruppi nell'insieme degli strati;

Si noti, in particolare, che altri test presenti in letteratura, come ad esempio il test basato sulle statistiche POSET proposto da Rosenbaum (1995), non permettono né l'analisi parziale di ciascun strato né l'analisi del contributo di ogni singolo strato. Tali test, infatti, forniscono solo la significatività complessiva dell'intero pattern su tutti gli strati considerati, ed inoltre hanno spesso poche proprietà statistiche a differenza della NPC che essendo una procedura condizionata ai dati, gode in generale di ottime proprietà statistiche.

*Osservazione 5.2:* Si vedrà successivamente nel capitolo 6 che si può estendere il principio che ad ogni variabile risposta considerata si applichi un solo test, o, più generalmente, si illustrerà come procedere in modo versatile alla scomposizione del problema di verifica.

---

<sup>5</sup> la metodologia di combinazione non parametrica (NPC) verrà discussa successivamente nel capitolo 6;

## 5.8.4 IL TEST DI PERMUTAZIONE SOTTO UN MODELLO LOGISTICO PER PROPENSITY SCORE

Si mostrerà come il test di permutazione di Fisher può essere esteso così da risultare applicabile agli studi osservazionali a condizione che

- (i) l'assegnazione ai trattamenti è fortemente ignorabile;
- (ii) il *propensity score* segue un modello logistico, cioè:

$$\log\{e(\mathbf{x})/[1 - e(\mathbf{x})]\} = \boldsymbol{\beta}^T \mathbf{f}(\mathbf{x}) \quad (5.4)$$

dove  $\boldsymbol{\beta}$  è un vettore di parametri non noto, e  $\mathbf{f}(\cdot)$  è un vettore di valori noto funzione di  $\mathbf{x}$ , ad esempio  $\mathbf{f}(\mathbf{x}) = (1, \mathbf{x})^T$

Siano  $\mathbf{F}$  ed  $\mathbf{X}$  le matrici in cui le  $n$  righe sono, rispettivamente, i valori di  $\mathbf{f}(\mathbf{x}_i)^T$  e  $\mathbf{x}_i^T$ ,  $i = 1, \dots, n$ . Si dimostra che (Cox 1970)  $\mathbf{z}^T \mathbf{F}$  è sufficiente per  $\boldsymbol{\beta}$  nella (5.4). Il test proposto è simile a un test di randomizzazione, ma con un parametro di disturbo  $\boldsymbol{\beta}$  che descrive il meccanismo di assegnazione al trattamento. Diversamente da un test di randomizzazione, questo test condizionato confronta la statistica test osservata  $t(\mathbf{z}, \mathbf{r})$  con il valore  $t(\mathbf{b}, \mathbf{r})$  che si avrebbe dovuto ottenere sotto l'ipotesi nulla con un assegnazione al trattamento diversa, indicata dal vettore binario  $\mathbf{b}$ , solo se  $\mathbf{b}$  è simile all'assegnazione al trattamento osservata, nel senso che

$$\mathbf{b}^T \mathbf{F} = \mathbf{z}^T \mathbf{F} \quad (5.5)$$

Ne deriva il seguente teorema (Rosenbaum, 1984):

*Teorema:* Si supponga che il *propensity score* segua il modello logistico (5.4):

- (a) La distribuzione condizionata delle assegnazioni al trattamento  $\mathbf{z}$  dato  $(\mathbf{z}^T, \mathbf{F}, \mathbf{X})$  è libera da parametri non noti e assegna la stessa probabilità a ogni vettore binario  $\mathbf{b}$  che soddisfa alla (5.5);
- (b) Sotto l'ipotesi nulla puntuale (5.2), se l'assegnazione al trattamento è fortemente ignorabile, allora la distribuzione condizionata della statistica test  $t(\mathbf{z}, \mathbf{r})$  dato  $\mathbf{r} = \mathbf{c}$  e  $(\mathbf{z}^T, \mathbf{F}, \mathbf{X})$  eguaglia la distribuzione di permutazione nota di  $t(\mathbf{z}, \mathbf{c})$  dato  $(\mathbf{z}^T, \mathbf{F}, \mathbf{X})$  determinata nella parte (a); cioè

$$\Pr\{t(\mathbf{z}, \mathbf{r}) \mid \mathbf{r} = \mathbf{c}, \mathbf{z}^T \mathbf{F}, \mathbf{X}\} = \Pr\{t(\mathbf{z}, \mathbf{c}) \mid \mathbf{z}^T \mathbf{F}, \mathbf{X}\} \quad \forall \text{ costante } \mathbf{c}$$

- (c) Se l'assegnazione al trattamento è fortemente ignorabile, e se, per ogni  $\mathbf{a}$  e  $\mathbf{c}$  fissato, l'insieme  $W(\mathbf{c}, \mathbf{a})$  soddisfa  $\Pr\{t(\mathbf{z}, \mathbf{c}) \in W(\mathbf{c}, \mathbf{a}) \mid \mathbf{z}^T \mathbf{F} = \mathbf{a}, \mathbf{X}\} = \alpha$  (rispettivamente  $\leq \alpha$ ), allora un test che rifiuta ogni volta che  $t(\mathbf{z}, \mathbf{r}) \in W(\mathbf{r}, \mathbf{z}^T \mathbf{F})$  ha il livello  $\alpha$  (rispettivamente  $\leq \alpha$ ) per tutti i valori del parametro non noto  $\boldsymbol{\beta}$ ;

## 5.9 NPC APPLICATA AD UNO STUDIO CLINICO E AD UN'INDAGINE EPIDEMIOLOGICA

Per enfatizzare la potenzialità, l'efficacia e l'utilità della metodologia di combinazione non parametrica si presentano ora due applicazioni rispettive ai due tipi di studi discussi nel presente capitolo, illustrate in Pesarin (2001). Ci si limiterà a esporre la descrizione del problema, presentandone la soluzione basata su NPC, ma omettendo i risultati dell'analisi.

### 5.9.1 STUDIO CLINICO SU UN FARMACO RESPIRATORIO

Si è considerato uno studio clinico su un farmaco potenzialmente respiratorio condotto in due centri (si veda Koch *et al.*, 1990). Uno degli scopi di tale studio era di dimostrare l'effetto di un dato farmaco respiratorio. A questo fine, i pazienti sono stati assegnati *a caso* ad uno dei due livelli del trattamento, il farmaco attivo  $A$  e un placebo  $P$ , dando origine a due gruppi di risposte. I pazienti sono stati esaminati in quattro visite sequenziali,  $X_1, \dots, X_4$ . La loro salute respiratoria è stata valutata come (1) molto scarsa, (2) scarsa, (3) discreta, (4) buona o (5) eccellente. Si è considerata una sola variabile di stratificazione, vale a dire l'età. Le risposte dei pazienti sono pertanto quattro variabili categoriali ordinate, corrispondenti alle quattro visite (e quindi misurazioni ripetute). Si è interessati a valutare se gli individui a cui è stato somministrato il farmaco si comportino meglio di coloro a cui è stato somministrato il placebo. Inoltre si vuole verificare  $A$  contro  $P$  con riferimento alla variabile di stratificazione scelta (che rappresenta l'unica covariata), l'età. Conformemente alle analisi proposte in Westfall e Young (1993) si è scelto di rendere dicotomica l'età in "giovane" ( $Y$ , età inferiore ai 32 anni) e "vecchio" ( $O$ , età maggiore o uguale ai 32 anni), affinché sia possibile condurre due analisi separate all'interno dei due strati.

Sia  $\mathbf{X} = \{X_{hsji}, i = 1, \dots, n_j; j = A, P; s = Y, O; h = 1, \dots, 4\}$  la notazione per l'intero insieme dei dati, dove  $X_{hsji}$  è la risposta dell' $i$ -esimo elemento del  $j$ -esimo gruppo nel  $s$ -esimo strato per l' $h$ -esima variabile, e  $n = 111$ . Inoltre, si denoti con  $X_{hsj}$  l' $h$ -esima variabile del  $j$ -esimo gruppo nel  $s$ -esimo strato. Le ipotesi multivariate globali possono essere scritte e corrispondentemente analizzate sia all'interno degli strati con rispetto alle variabili:

$$H_0 : \{\cap_s [\cap_h (X_{hAs}^d = X_{hPs}^d)]\} \text{ contro } H_1 : \{\cup_s [\cup_h (X_{hAs}^d > X_{hPs}^d)]\}$$

sia entro le variabili, con rispetto agli strati:

$$H_0 : \{\cap_h [\cap_s (X_{hAs}^d = X_{hPs}^d)]\} \text{ contro } H_1 : \{\cup_h [\cup_s (X_{hAs}^d > X_{hPs}^d)]\}$$

dove viene enfatizzata la scomposizione in un insieme di sotto-ipotesi appropriate. Si noti che il problema è formalmente un problema di *goodness of fit* multivariato per variabili categoriali ordinate sotto restrizioni d'ordine. Tramite l'approccio basato sul rapporto di verosimiglianza, sembra che in letteratura non venga fornita alcuna soluzione soddisfacente quando vi sono più di due misurazioni ripetute, o in generale, più di due variabili.

Si vuole sottolineare dunque che la discussione relativa alla “teoria elaborata” presentata nel paragrafo 5.7 non è circoscritta ai soli studi osservazionali, in quanto in problemi complessi è comune definire un pattern multivariato, come si è visto nel presente esempio.

### 5.9.2 UN'INDAGINE EPIDEMIOLOGICA : IL PROGETTO SETIG

Nel triennio '95 – '97 è stato pianificato e condotto il progetto SETIG – Sorveglianza Epidemiologica dei Trattamenti delle Infezioni Gravi – con l'obiettivo di approfondire il problema delle infezioni trattate in ambito ospedaliero, con riguardo al confronto tra i profili di risposta di pazienti con sepsi sottoposti ad un trattamento mirato (con antibiotici specifici per il patogeno individuato da un'analisi microbiologica) e di pazienti trattati in modo empirico (con antibiotici ad ampio spettro).

I fattori risultati confondenti da una serie di analisi esplorative sono stati: l'età del paziente, il numero dei fattori predisponenti l'infezione e la presenza di tumore, per quanto riguarda le caratteristiche individuali del paziente; il reparto di ricovero e l'area geografica del centro partecipante, come variabili riflettenti la componente culturale, strutturale e organizzativa della gestione del problema infezione. I pazienti sono stati quindi stratificati secondo classi di valori definiti dal *propensity score*. Per il calcolo del *propensity score* è stata applicata una regressione logistica con il tipo di terapia come variabile dipendente e le variabili: età, numero di fattori predisponenti, presenza di tumore, tipo di reparto, area geografica come variabili indipendenti. La probabilità di ricevere un trattamento empirico stimata da questo modello è stata quindi utilizzata per creare quattro strati omogenei per *propensity score* (mediante il calcolo dei quartili della distribuzione del *propensity score*).

Il test di permutazione multivariato e multistrato basato sulla NPC è stato applicato considerando quattro variabili risposta che componevano il pattern causale: (*U*) durata del trattamento, che assumeva valori 0, 1 (0 = ‘7-15 giorni del trattamento’, 1 = ‘>15

giorni'); ( $D$ ) decesso a valori 0, 1 (0 = 'vivo', 1 = 'morto'); ( $L$ ) degenza (numero di giorni di ricovero); ( $R$ ) risoluzione clinica dell'infezione, a valori 0, 1 (0 = 'non risolta', 1 = 'risolta'). Il sistema multidimensionale di ipotesi globale che definisce il pattern causale può essere scritto, e corrispondentemente analizzato, sia all'interno degli strati, con rispetto alle variabili

$$H_0 : \cap_s \{(D_{1s}^d = D_{2s}^d) \cap (R_{1s}^d = R_{2s}^d) \cap (U_{1s}^d = U_{2s}^d) \cap (L_{1s}^d = L_{2s}^d)\}$$

contro

$$H_1 : \cup_s \{(D_{1s}^d > D_{2s}^d) \cup (R_{1s}^d < R_{2s}^d) \cup (U_{1s}^d > U_{2s}^d) \cup (L_{1s}^d > L_{2s}^d)\}$$

oppure entro le variabili, con rispetto agli strati

$$H_0 : \{[\cap_s (D_{1s}^d = D_{2s}^d)] \cap [\cap_s (R_{1s}^d = R_{2s}^d)] \cap [\cap_s (U_{1s}^d = U_{2s}^d)] \cap [\cap_s (L_{1s}^d = L_{2s}^d)]\}$$

contro

$$H_1 : \{[\cup_s (D_{1s}^d > D_{2s}^d)] \cup [\cup_s (R_{1s}^d < R_{2s}^d)] \cup [\cup_s (U_{1s}^d > U_{2s}^d)] \cup [\cup_s (L_{1s}^d > L_{2s}^d)]\}$$

Si noti che questo problema presenta, in ciascun strato, tre variabili binarie e una quantitativa; inoltre, tutte le sotto-alternative sono unidirezionali, tre con una direzione positiva e una con direzione negativa. In aggiunta, le variabili  $U$  e  $L$  presentano valori mancanti che potrebbero essere persi non a caso. In generale, il trattamento dei dati mancanti risulta possibile soltanto quando vale l'assunto che i dati mancanti siano mancanti completamente a caso, ossia che la probabilità che un dato venga a mancare non risulti affetta dal livello del trattamento. Nell'ambito dei metodi di permutazione e di combinazione NPC è possibile trattare i dati mancanti anche quando tale assunto non risulta soddisfatto (si veda Pesarin, 2001).

# Capitolo 6

## I Test Multi-Aspetto

### 6.1 INTRODUZIONE

Il “*multiple testing*” si riferisce ad una qualsiasi situazione in cui una collezione di statistiche test viene valutata formalmente o informalmente. Questa situazione si applica virtualmente a tutte le analisi dei dati nelle quali vengono considerate delle ipotesi; è infatti molto inusuale che su un insieme di dati venga effettuato un solo test. Vi è una crescente tendenza ad analizzare insiemi di dati da ogni possibile angolo, formulando e provando miriadi di ipotesi. Perché effettuare una verifica dei dati così esauriente?

- (1) la disponibilità di software statistico moderno ha reso abbastanza semplice l'esecuzione di verifiche multiple. Una volta che l'insieme dei dati è *disponibile*, di solito si impiega poco tempo per formulare e verificare ipotesi. D'altra parte, *l'acquisizione* e il *pretrattamento* dei dati risulta laborioso e caro. Quindi, non appena l'insieme dei dati è sul computer, ne viene analizzato ogni singolo aspetto;
- (2) si possono utilizzare diversi test statistici per fornire un'evidenza confermativa dell'ipotesi verificata: se c'è qualche dubbio a proposito della validità di una particolare procedura di verifica, si considera prudente eseguire il test utilizzando metodi alternativi;
- (3) nel verificare l'ipotesi nulla contro tipi qualitativamente differenti di deviazioni da essa, si possono progettare statistiche appropriate per le diverse deviazioni: ad esempio, una distribuzione può deviare dall'ipotesi nulla di normalità per asimmetria, per curtosi oppure per entrambe, e quindi si possono utilizzare statistiche che verifichino ciascuno di questi aspetti<sup>1</sup>;

Lo svantaggio principale della verifica multipla è la probabilità di dichiarare “false significatività”, cioè associazioni statisticamente significative quando nella realtà non esistono. Una caratteristica negativa connessa a ciò è che è molto facile *esagerare*<sup>2</sup> l'evidenza di una particolare associazione se viene scelto il test statistico che meglio sostiene una data ipotesi.

---

<sup>1</sup> Cox, 1977;

<sup>2</sup> “*overstate*”, Westfall e Young, 1993;

Questi problemi sono particolarmente perniciosi quando si considera che le analisi statistiche pubblicate spesso presentano solo i risultati di alcuni dei molti test considerati o calcolati. Non vi è di solito alcuna indicazione sul numero delle indagini che hanno preceduto il rapporto finale. L'uso di eccessive verifiche di ipotesi statistiche viene spesso chiamato “*data snooping*” oppure “*data dredging*” in un tentativo di disonorare gli investigatori. I risultati di una tale manipolazione dei dati sono comunemente manufatti delle tecniche impiegate e dovrebbero essere scartati come “risultati spuri”, ma spesso non lo sono. Di conseguenza vi dovrebbe essere una qualche garanzia che l'effetto “statisticamente significativo” riportato non sia solo un prodotto del *data dredging*, garanzia che può essere ottenuta mediante il controllo del “*maximum familywise error rate*”<sup>3</sup> (spesso abbreviato FWE), cioè la probabilità massima che una o più ipotesi nulle siano rifiutate erroneamente.

*Osservazione 6.1:* Per le analisi “confermative” emerge un problema: dal momento che i test sono fortemente correlati, tenderanno a concordare anche quando la conclusione è un risultato falsamente positivo (Westfall e Young, 1993).

## 6.2 IL “MAXIMUM” TEST

Cox e Hinkley (1974) discutono il caso che si presenta quando un certo numero di test di significatività vengono applicati per verificare la stessa ipotesi nulla: la consuetudine deliberata di applicare numerosi test nella speranza di trovare un risultato “significativo” non viene raccomandata e ciò deve essere ravvisato se si vuole evitare di trarre conclusioni ingannevoli.

Si supponga che siano stati utilizzati  $k$  test e che venga considerato il più significativo tra di essi. Ci sono certamente altri modi per combinare i risultati di più test, ma si supponga di voler rappresentare una situazione in cui vi siano un certo numero di differenti tipi di deviazioni da  $H_0$  e quella infine esaminata sia la più significativa tra di esse. Questa procedura viene denominata “*maximum*” test<sup>4</sup>. Siano  $P_1, \dots, P_k$  i livelli di significatività dei  $k$  test; sia  $Q$  una statistica tale che:

$$Q = \min(P_1, \dots, P_k)$$

Nel caso continuo,  $P_1, \dots, P_k$  sono uniformemente distribuiti sotto  $H_0$ , pertanto

<sup>3</sup> questa quantità viene anche chiamata “*maximum overall type I error rate*” o “*maximum experimentwise error rate*”, Westfall e Wolfinger, 2000;

<sup>4</sup> Cox, 1977;

$$\Pr\{Q > q_{obs}; H_0\} = \Pr\{P_j > q_{obs}, j = 1, \dots, k; H_0\}$$

quindi il livello di significatività del test  $Q$  risulta

$$\Pr\{Q \leq q_{obs}; H_0\} = 1 - \Pr\{P_j > q_{obs}, j = 1, \dots, k; H_0\}$$

Se i  $k$  test sono indipendenti

$$\Pr\{Q \leq q_{obs}; H_0\} = 1 - \prod_{j=1}^k \Pr\{P_j > q_{obs}\} = 1 - (1 - q_{obs})^k$$

Se al contrario i  $k$  test non sono indipendenti (solitamente si presenterà dipendenza positiva poiché sono calcolati sui medesimi dati), si dovrebbe valutare  $\Pr\{Q \leq q_{obs}; H_0\}$  dalla distribuzione congiunta di  $P_1, \dots, P_k$ ; ciò non sarà sempre possibile, pertanto si può ottenere un limite superiore del livello di significatività nel seguente modo:

Per eventi arbitrari  $A_1, \dots, A_k$ , risulta  $A_1 \cap \dots \cap A_k$  complementare di  $\bar{A}_1 \cup \dots \cup \bar{A}_k$

$$\Pr\{A_1 \cap \dots \cap A_k\} = 1 - \Pr\{\bar{A}_1 \cup \dots \cup \bar{A}_k\}$$

$$1 - \Pr\{A_1 \cap \dots \cap A_k\} = \Pr\{\bar{A}_1 \cup \dots \cup \bar{A}_k\} \leq \sum_{j=1}^k \Pr\{\bar{A}_j\} \quad (5)$$

$$\Rightarrow \Pr\{Q \leq q_{obs}; H_0\} \leq \sum_{j=1}^k \Pr\{P_j \leq q_{obs}\} = k \cdot q_{obs}$$

cioè,  $k \cdot q_{obs}$  rappresenta il limite superiore del livello di significatività.

Nel problema di verifica dell'ipotesi nulla contro differenti tipi di deviazioni, l'approccio matematicamente più semplice è quello di progettare *a priori* statistiche separate per i diversi tipi di deviazioni e quindi combinarle in una funzione sensibile alle deviazioni in alcuni o in tutti gli aspetti considerati.

Si presume che la procedura “*maximum*” risulti sensibile nel caso di deviazioni “pure” dall'ipotesi nulla (quando vi è un'unica deviazione in una determinata “direzione” o aspetto), mentre risulterà meno potente nel caso di deviazioni “miste” (quando le deviazioni, anche se relativamente piccole, si presentano simultaneamente in diverse “direzioni”)<sup>4</sup>.

<sup>4</sup> L'uguaglianza è verificata nel caso di una successione di eventi  $\bar{A}_1, \dots, \bar{A}_k$  a due a due incompatibili:  $\bar{A}_i \cap \bar{A}_j = \emptyset, i \neq j$ ;



### 6.3 APPROCCIO MULTI-ASPECT

Si riprenda in considerazione l'esperimento storico sul grado dello sviluppo discusso nel capitolo 5 e si ragioni sulle seguenti considerazioni di sir Ronald Fisher: «...se lo scopo dell'esperimento fosse stato, non di accertare se l'altezza media delle piante a fecondazione incrociata fosse o no più grande di quelle auto-fecondate, ma fosse stato invece di accertare se le differenze in altezza tra piante crociate o auto-fecondate fossero distribuite normalmente o in una distribuzione asimmetrica, allora si poteva escogitare una prova valida appropriata a questo punto. (...)»<sup>6</sup> Questa prova discernerebbe come statisticamente significativa delle serie di risultati sperimentali completamente differenti da quelle scelte con la prova *t*. Essa non è affatto un sostituto di questa prova o qualcosa di adatto a compiere le medesime funzioni. Essa si propone di rispondere a un differente quesito sebbene in ambedue i casi si risponde al quesito selezionando un gruppo di possibili risultati sperimentali giudicati tali da contraddire la stessa ipotesi zero. Esse si possono propriamente pensare come capaci di provare differenti aspetti della stessa ipotesi. L'ipotesi provata in ambedue i casi ammette che la distribuzione delle differenze di altezza è centrata in zero ed è normale nella forma. Una delle prove è appropriata nel caso in cui il nostro speciale interesse è rivolto alla possibilità che essa non è centrata a zero. In questo caso la questione della normalità è, come si è dimostrato, di importanza del tutto trascurabile. L'altra prova è appropriata nel caso in cui il nostro interesse è rivolto alla possibilità che la distribuzione sia sghemba, cioè asimmetrica intorno alla sua media e in tal caso il valore della media è completamente trascurabile.

La nozione che differenti test di significatività sono adatti a provare differenti aspetti della stessa ipotesi nulla non presenta difficoltà per i ricercatori nella sperimentazione pratica, ma è stata l'occasione di molte discussioni teoretiche fra gli statistici. La ragione che giustifica questi diversi punti di vista è forse il fatto che lo sperimentatore pensa in termini di valori di osservazione, ed esso è consapevole che quella discrepanza che a lui interessa e che ritiene possa essere statisticamente significativa, ancor prima cerchi quale test di significatività, se pur esiste, sia vantaggiosamente appropriato alle sue necessità. Egli perciò di solito non si preoccupa del quesito: "A quale aspetto delle osservazioni deve essere applicato un test di significatività?" Tale quesito, quando la risposta non è già conosciuta, può essere vantaggiosamente discusso solo quando lo sperimentatore si prospetta non una sola ipotesi nulla, ma una classe di tali ipotesi della significatività delle deviazioni da ognuna delle quali egli è ugualmente interessato.

Non si deve, però, pensare che un tale fondamento teorico sia normale condizione della sperimentazione o che sia necessaria per un adeguato uso delle prove di significatività».

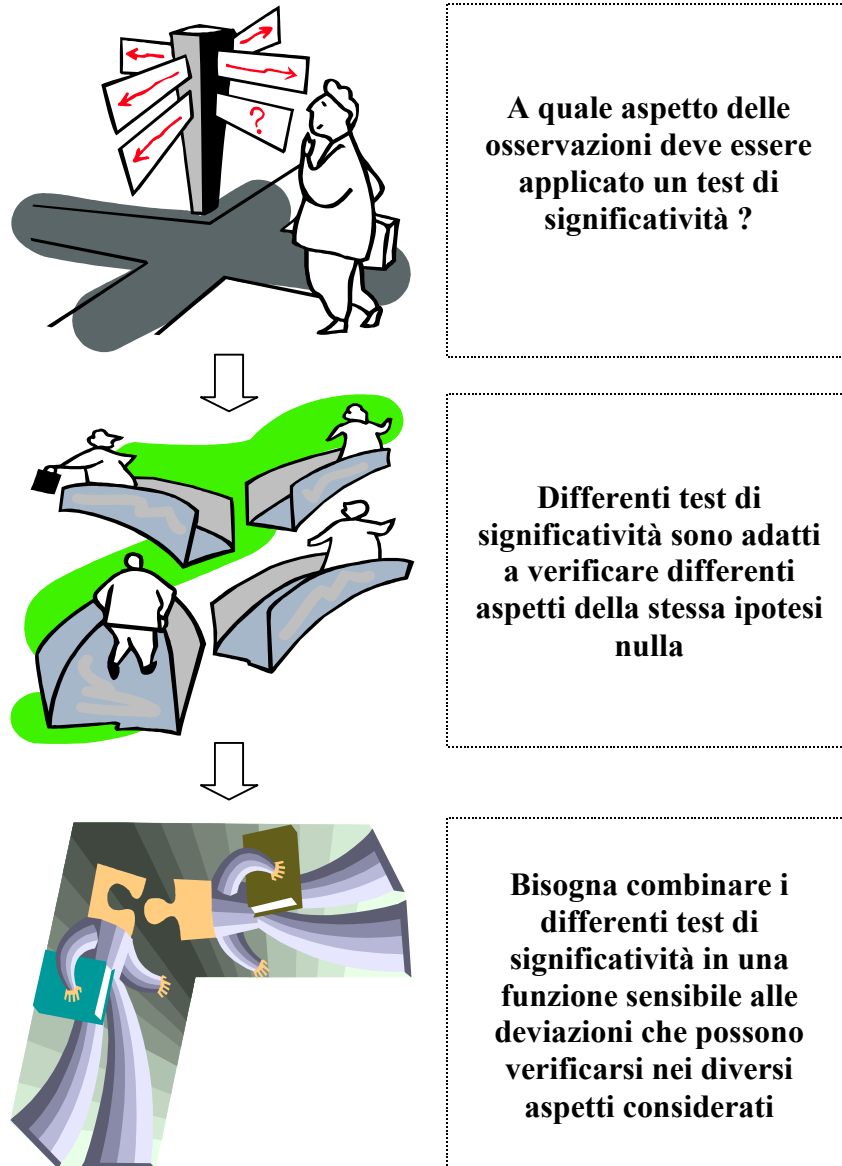
(Fisher, 1935)

---

<sup>6</sup> Fisher propone di «calcolare la somma dei cubi di queste differenze, tenendo conto dei segni e il rapporto tra quest'ultima somma e la prima (la somma dei quadrati delle deviazioni di queste differenze) elevata alla potenza di 3/2»;

Riflettendo sulle considerazioni di Fisher, nella figura sottostante viene illustrata una metafora del problema: gli individui rappresentano i test di significatività mentre le diverse direzioni rappresentano i differenti aspetti della stessa ipotesi nulla.

Figura 6.1:



Si consideri un problema univariato per il quale è di interesse verificare una certa ipotesi nulla: se è possibile considerare più d'un aspetto per tale ipotesi, ovvero se si possono presentare deviazioni qualitativamente differenti da essa, allora è possibile scomporre l'ipotesi nulla iniziale in un'insieme di sotto-ipotesi, ognuna peculiare per ciascun aspetto. Quindi è opportuno, in un primo passo, progettare i test parziali relativi agli aspetti considerati, e, in un secondo passo, valutarli congiuntamente per verificare l'ipotesi nulla iniziale, che rappresenta il vero obiettivo dell'inferenza.

Si consideri ora una situazione più complessa nella quale vengono considerate  $q$  variabili per verificare un certa ipotesi nulla globale: è comune che il problema venga scomposto in

$q$  sotto-problemi, ognuno relativo a ciascuna variabile considerata. Si potrebbe considerare ciascuna variabile componente come un “aspetto” dello stesso insieme di dati: in questo modo sarebbe possibile interpretare il problema multivariato in conformità alla logica Multi-Aspetto o *Multi-Aspect*.

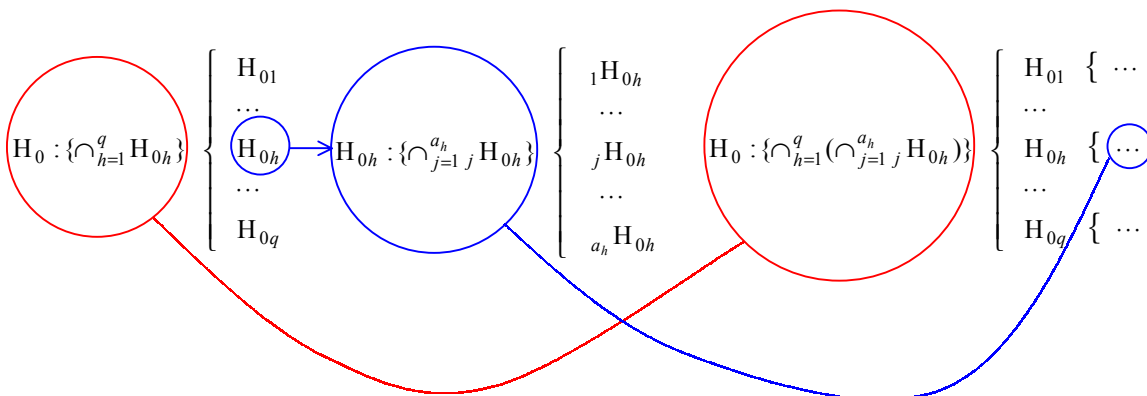
In realtà, sia l’approccio multi-aspect “ristretto” univariato sia quello “esteso” multivariato aderiscono come casi particolari allo stesso procedimento: il metodo di combinazione nonparametrica di un numero finito di test di permutazione dipendenti (Pesarin, 2001).

Per comprendere ciò, si consideri la versatilità della metodologia di combinazione nonparametrica: essa afferma che *non necessariamente* il numero dei sotto-problemi debba essere uguale alla dimensionalità delle risposte.

Infatti, in un problema  $q$ -variato, l’ $h$ -esima sotto-ipotesi ( $h = 1, \dots, q$ ) potrebbe essere ulteriormente scomposta in  $a_h$  aspetti se le *side-assumptions* per il problema permettono di considerare più d’un aspetto per tale ipotesi (tabella 6.1):

**Tabella 6.1:** Scomposizione del problema secondo la procedura di combinazione non parametrica

problema $q$ -variato al quale corrispondono $q$ sotto-ipotesi	problema univariato al quale corrispondono $a_h$ aspetti	problema $q$ -variato al quale corrispondono $k = \sum_{h=1}^q a_h$ sotto-ipotesi
--	--	---



In generale, quindi, ipotizzando delle *side-assumptions* appropriate per il problema, la metodologia di combinazione nonparametrica permette di scomporre l’ipotesi nulla globale  $H_0$  in un’insieme finito di sotto-ipotesi  $H_{0i}, i = 1, \dots, k$ . D’ora in poi, ci si riferirà all’approccio multi-aspect “ristretto”, denominandolo multi-aspect senza ulteriori specificazioni.

Altre caratteristiche della metodologia di combinazione nonparametrica che verranno “ereditate” dall’approccio multi-aspect sono le seguenti:

- (1) Ciascuna statistica test risulta informativa per un aspetto rilevante dell’analisi, e quindi, utile per il medesimo; tuttavia, può accadere che tutte insieme siano

congiuntamente informative, e quindi emerge la necessità di prendere in considerazione tutte le informazioni disponibili in un test di ordine superiore;

- (2) Risulta impossibile assumere la completa indipendenza dei  $a$  test parziali in quanto sono funzioni degli stessi dati (tabella 6.2); inoltre, le relazioni di dipendenza sottostanti sono raramente conosciute ad eccezione di situazioni molto semplici, e anche se note spesso risulta troppo difficile far fronte ad esse, quindi la combinazione deve essere fatta in maniera non parametrica, con special riguardo alle relazioni di dipendenza sottostanti, che vengono “catturate” dalla procedura di combinazione senza la necessità di modellarle;

**Tabella 6.2:** Approccio multi-aspect al problema univariato a due campioni: statistiche test dipendenti e test combinato

<i>ipotesi nulla</i>	<i>insieme dei dati</i>	<i>statistiche test</i>	<i>test combinato</i>
$H_0 : \{X_1 = X_2\}^d$ $\swarrow \quad \downarrow \quad \searrow$ $A_1 \dots A_j \dots A_a$ $\downarrow \quad \downarrow \quad \downarrow$ $H_{01} \dots H_{0j} \dots H_{0a}$	$\mathbf{X} = \{X_{ji}, i = 1, \dots, n_j, j = 1, 2\}$	$\left. \begin{array}{l} T_1 : \Omega^n \rightarrow \mathfrak{R}^1 \text{ per } H_{01} \\ \dots \\ T_j : \Omega^n \rightarrow \mathfrak{R}^1 \text{ per } H_{0j} \\ \dots \\ T_a : \Omega^n \rightarrow \mathfrak{R}^1 \text{ per } H_{0a} \end{array} \right\}$	$\left. \begin{array}{l} T^* = \psi(\lambda_1, \dots, \lambda_a) \\ \psi : (0, 1)^a \rightarrow \mathfrak{R}^1 \end{array} \right\}$

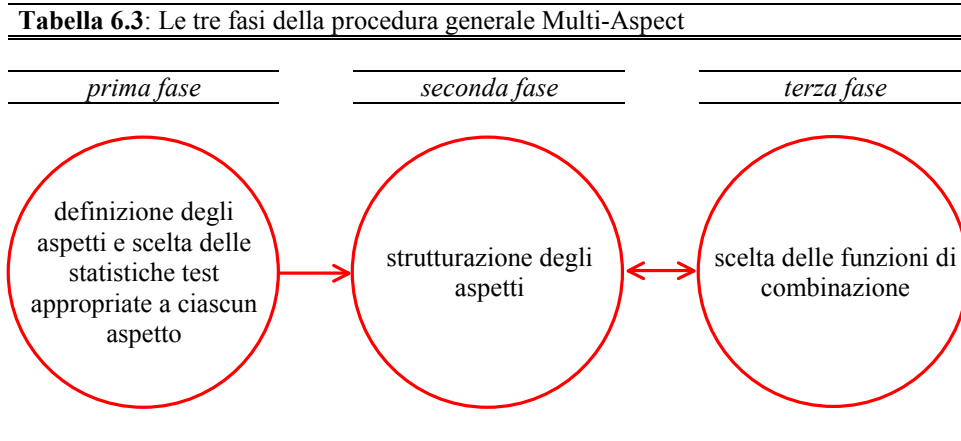
- (3) Soluzioni conservative possono essere ottenute attraverso la disuguaglianza di Bonferroni; la conservatività risulta spesso inaccettabile sia per finalità teoriche che pratiche. Si vuole far notare che le procedure di confronto multiplo hanno come punto di partenza il test globale e cercano i test significativi sui contrasti parziali; al contrario, le procedure di combinazione partono con un’insieme di test parziali, ciascuno appropriato per un aspetto, e mirano ad un’analisi congiunta che conduce all’inferenza globale;

Nel paragrafi successivi verrà illustrata la procedura generale multi-aspect e la metodologia di combinazione nonparametrica di un numero finito di test di permutazione dipendenti; d’ora in poi, per chiarezza di esposizione, le notazioni saranno le stesse utilizzate in Pesarin (2001).

<sup>7</sup> la funzione di combinazione  $\psi$  verrà definita nei paragrafi 6.5.2 e 6.7;

## 6.4 PROCEDURA GENERALE MULTI-ASPECT

Nella tabella sottostante (tabella 6.3) viene schematizzata la procedura generale a tre fasi:



Nella *prima fase*, si procede alla definizione degli aspetti da considerare al fine da verificare una certa ipotesi nulla; in seguito si definiscono le statistiche test appropriate a ciascun aspetto e le relative sotto-ipotesi; il principio che ad ogni aspetto corrisponda una sola statistica test può essere esteso affinché ad uno stesso aspetto possano corrispondere più statistiche test (tabella 6.3a);

**Tabella 6.3a:** Prima fase

<i>ipotesi nulla</i>	<i>aspetti considerati</i>	<i>statistiche test</i>	<i>sotto-ipotesi</i>
$H_0$	$A_1$	$T_1$	$\Rightarrow H_{01}$
	$\dots$	$T_2$	$\Rightarrow H_{02}$
	$\dots$	$\dots$	$\dots$
$H_0$	$A_j$	$T_i$	$\Rightarrow H_{0i}$
	$\dots$	$\dots$	$\dots$
$H_0$	$A_a$	$\dots$	$\dots$
	$\dots$	$T_k$	$\Rightarrow H_{0k}$

Nella *seconda fase*, si procede alla strutturazione degli aspetti: nella tabella 6.3b sono illustrati due possibili esempi: il primo struttura “direttamente” tutti gli aspetti considerati, dando ad ognuno di essi lo stesso “peso” nella combinazione finale; il secondo, invece,

struttura due insiemi disgiunti e complementari di aspetti, dando origine a due macro-aspetti, a cui è assegnata la stessa importanza nella combinazione conclusiva;

**Tabella 6.3b:** Seconda fase

<i>aspetti</i>	<i>sotto-ipotesi associate</i>	<i>esempi di strutturazioni</i>	<i>ipotesi nulla associate</i>	
$A_1$	$H_{A_1} : \{H_{01} \cap H_{02} \cap \dots\}$		$H_0 : \{(H_{01} \cap H_{02} \cap \dots) \cap$ $(\dots \cap H_{0i} \cap \dots) \cap$ $(\dots \cap H_{0k})\}$	
...	...			
$A_j$	$H_{A_j} : \{\dots \cap H_{0i} \cap \dots\}$		$[A_1 \cap \dots \cap A_j]$	$H_0 : \{[(H_{01} \cap H_{02} \cap \dots) \cap$ $(\dots \cap H_{0i} \cap \dots)]$ $\cap$ $[(\dots) \cap \dots \cap (\dots \cap H_{0k})]\}$
...	...		$\cap$	
$A_a$	$H_{A_a} : \{\dots \cap H_{0k}\}$		$[A_{j+1} \cap \dots \cap A_a]$	



**Osservazione 6.2:** Si vuole enfatizzare la versatilità della metodologia di combinazione nonparametrica: è infatti possibile qualsiasi tipo di strutturazione, e quindi qualsiasi tipo di gerarchia di aspetti, che può essere utile a scopi interpretativi; inoltre, la strutturazione permette di “pesare” gli aspetti nella funzione di combinazione finale;

Nella *terza fase*, si effettua la scelta delle funzioni di combinazione da utilizzare; si può procedere in due passi: (i) all’interno di ogni aspetto considerato; (ii) tra gli aspetti considerati, in base alla strutturazione realizzata nella seconda fase; (rispettivamente le tabelle 6.3c-i e 6.3c-ii)

**Tabella 6.3c-i:** Terza fase, primo passo

<i>aspetti considerati</i>	<i>statistiche test</i>	<i>p-values dei test</i>	<i>sotto-ipotesi dell’aspetto</i>	<i>test combinato associato all’aspetto</i>	<i>p-values dell’aspetto</i>
$A_1$	$(T_1, T_2, \dots)$	$(\lambda_1, \lambda_2, \dots)$	$(H_{01} \cap H_{02} \cap \dots)$	$T_{A_1}'' = \psi_l(\lambda_1, \lambda_2, \dots)$	$\lambda_{A_1}''$
...	...	...	...	...	...
$A_j$	$(\dots, T_i, \dots)$	$(\dots, \lambda_i, \dots)$	$(\dots \cap H_{0i} \cap \dots)$	$T_{A_j}'' = \psi_l(\dots, \lambda_i, \dots)$	$\lambda_{A_j}''$
...	...	...	...	...	...
$A_a$	$(\dots, T_k)$	$(\dots, \lambda_k)$	$(\dots \cap H_{0k})$	$T_{A_a}'' = \psi_l(\dots, \lambda_k)$	$\lambda_{A_a}''$

**Tabella 6.3c-ii:** Terza fase, secondo passo

	<i>esempi di struttura</i>	<i>combinazione intermedia</i>	<i>p-values</i>	<i>combinazione globale</i>	<i>p-value globale</i>
$A_1$				$T_{MA}''' = \psi_l(\lambda_{A_1}''', \dots, \lambda_{A_a}''')$	$\lambda_{MA}'''$
...		$T_{A_1 \dots A_j}''' = \psi_l(\lambda_{A_1}''', \dots, \lambda_{A_j}''')$	$\lambda_{A_1 \dots A_j}'''$		
$A_j$					
...		$T_{A_{j+1} \dots A_a}''' = \psi_l(\lambda_{A_{j+1}}''', \dots, \lambda_{A_a}''')$	$\lambda_{A_{j+1} \dots A_a}'''$	$T_{MA}''' = \psi_l(\lambda_{A_1 \dots A_j}''', \lambda_{A_{j+1} \dots A_a}''')$	$\lambda_{MA}'''$
$A_a$					

dove  $\psi_l$  è una funzione di combinazione tale che  $l \in C$  (ad esempio,  $l = \text{Fisher}$ ; la scelta viene effettuata per ogni combinazione); in generale, l'ordine del test finale dipenderà dal tipo di strutturazione predisposta.

*Osservazione 6.3:* Si suggerisce di progettare la seconda e la terza fase congiuntamente (in tal senso nella tabella 6.3 si è inserito il simbolo “ $\leftrightarrow$ ”), in quanto la strutturazione degli aspetti e la scelta delle funzioni di combinazione sono due decisioni interdipendenti che vanno considerate contemporaneamente.

Si consideri ora una forma di combinazione neutra, denominata procedura di combinazione iterata: essa è una procedura multi-aspect poiché applica ai medesimi test parziali diverse funzioni di combinazione (tabella 6.4);

**Tabella 6.4:** Procedura iterata e multi-aspect

<i>insieme di p-values di ordine o</i>	<i>s test combinati di ordine o+1</i>	<i>p-values di ordine o+1</i>	<i>test combinato di ordine o+2</i>	<i>p-value di ordine o+2</i>
$(\dots, \lambda_h^o, \dots)$	$\left\{ \begin{array}{l} T^{o+1} = \psi_{l_1}(\dots, \lambda_h^o, \dots) \\ \dots \\ T^{o+1} = \psi_{l_s}(\dots, \lambda_h^o, \dots) \end{array} \right.$	$\left\{ \begin{array}{l} \lambda_{l_1}^{o+1} \\ \dots \\ \lambda_{l_s}^{o+1} \end{array} \right.$	$T^{o+2} = \psi_l(\lambda_{l_1}^{o+1}, \dots, \lambda_{l_s}^{o+1})$	$\lambda_l^{o+2}$

Infatti, differenti funzioni di combinazione, a causa delle differenti regioni di rifiuto<sup>8</sup>, possono fornire  $p$ -values globali differenti, sebbene per la loro consistenza<sup>9</sup>, siano asintoticamente equivalenti nell'alternativa. In modo da ridurre questa influenza, si può iterare la procedura di combinazione applicando ai medesimi test parziali di ordine  $o$  più

<sup>8</sup> si veda la figura 6.2, paragrafo 6.7;

<sup>9</sup> si veda paragrafo 6.5.5;

d'una funzione di combinazione  $\psi_{l_1}, \dots, \psi_{l_s}, 2 \leq l_s$ , e quindi combinare i  $p$ -values risultanti di ordine  $o+1$  ( $\lambda_{l_1}^{o+1}, \dots, \lambda_{l_s}^{o+1}$ ) in una combinazione di ordine  $o+2$  per mezzo di una funzione di combinazione  $\psi_l(\lambda_{l_1}^{o+1}, \dots, \lambda_{l_s}^{o+1})$ . Da una serie di esperimenti Monte Carlo, a condizione che le funzioni di combinazione abbiano differenti regioni di rifiuto, si è ottenuto che il  $p$ -value di ordine  $o+2$  dovrebbe essere quasi invariante rispetto alla scelta di  $\psi_l$  all'interno della classe  $C$ . Inoltre, per la versatilità della procedura, è possibile progettare ulteriori iterazioni.

Con riferimento alla procedura multi-aspect, la combinazione neutra può essere utilizzata nei passi (i) e (ii) della terza fase sopra descritta; questa forma di combinazione può essere considerata *robusta* rispetto alla funzione di combinazione.

*Osservazione 6.4:* La scelta delle statistiche test e delle funzioni di combinazione verrà discussa in seguito rispettivamente ai paragrafi 6.6 e 6.7;

*Osservazione 6.5:* La procedura multi-aspect nella sua forma generale ha inizio dalla definizione degli aspetti: ciò può sembrare un atto arbitrario, e si potrebbe semplicemente considerare un insieme di statistiche test e le relative sotto-ipotesi ritenute appropriate<sup>10</sup>.

Si vuole enfatizzare che la definizione degli aspetti e la conseguente strutturazione forniscono un'indicazione su come procedere nel processo:

- la definizione degli aspetti risulta utile nell'interpretazione dell'ipotesi nulla e nella scelta delle statistiche test associate ad ogni aspetto;
- se ad un aspetto è associato un insieme di statistiche test eterogenee, esso può venir considerato *robusto* rispetto alla scelta di tali statistiche, poiché nel passo successivo verranno combinate, “pesando” inoltre ogni aspetto iniziale allo stesso modo, e non in funzione del numero delle statistiche test considerate;
- la strutturazione degli aspetti permette di decidere *quando* e *in che modo* gli aspetti debbano essere combinati, dando origine ad una gerarchia di aspetti che ne determina l'importanza e che può essere utile a scopi interpretativi;
- Si vuole enfatizzare che, la verifica parziale o *componetwise testing*, possibilmente dopo una correzione dei  $p$ -value dovuta alla molteplicità, può essere utile per verificare le inferenze marginali; quindi, da un lato, i test parziali forniscono l'informazione marginale per ciascuna sotto-ipotesi specifica, dall'altro, congiuntamente, forniscono l'informazione sull'ipotesi globale  $H_0$  (per una discussione maggiormente approfondita si veda il capitolo 7, paragrafo 7.10);

---

<sup>10</sup> in realtà questa situazione è un caso particolare della procedura, che si presenta quando si associa ad ogni singolo aspetto una sola statistica test e la relativa sotto-ipotesi;



## 6.5 MULTI-ASPECT COME CASO PARTICOLARE DELLA METODOLOGIA DI COMBINAZIONE NONPARAMETRICA PER TEST DIPENDENTI

Si vogliono ora introdurre le assunzioni principali riguardanti l'insieme dei test parziali, le proprietà desiderabili delle funzioni di combinazione e l'algoritmo per la combinazione non parametrica. Inoltre, sulla base di queste assunzioni e proprietà, verrà dimostrata la non distorsione e la consistenza del test combinato.

### 6.5.1 ASSUNZIONI SUI TEST PARZIALI

Specifichiamo le assunzioni riguardanti l'insieme dei test parziali  $\mathbf{T} = \{T_i, i = 1, \dots, k\}$  sufficienti per la combinazione nonparametrica:

- (A.1) Tutti i test parziali di permutazione  $T_i$  sono marginalmente non distorti e significativi per valori elevati, tali da essere stocasticamente superiori in  $H_1$  rispetto ad  $H_0$ , in senso sia condizionato che incondizionato;
- (A.2) Tutti i test parziali di permutazione sono consistenti, cioè  $\Pr\{T_i \geq T_{i\alpha}\} \rightarrow 1, \forall \alpha > 0, i = 1, \dots, k$  quando  $n$  tende all'infinito, dove  $T_{i\alpha}$ , che è assunto essere finito, rappresenta il valore critico di  $T_i$  al livello  $\alpha$ ;

*Osservazione 6.6:* L'assunzione (A.1) formalmente implica che:

$$\begin{aligned} \Pr\{T_i \leq T_{i\alpha} \mid \mathbf{X}, H_{1i}\} &\geq \alpha, \quad \forall \alpha, i = 1, \dots, k \text{ e} \\ \Pr\{T_i \leq z \mid \mathbf{X}, H_{0i}\} &= \Pr\{T_i \leq z \mid \mathbf{X}, H_{0i} \cap H_i^\dagger\} \\ &\geq \Pr\{T_i \leq z \mid \mathbf{X}, H_{1i}\} = \Pr\{T_i \leq z \mid \mathbf{X}, H_{1i} \cap H_i^\dagger\} \end{aligned} \quad i = 1, \dots, k, \forall z \in \mathfrak{R}^1$$

dove l'irrelevanza rispetto all'insieme complementare di ipotesi  $H_i^\dagger \{\cup_{j \neq i} (H_{0j} \cup H_{1j})\}$  significa che non ha importanza quale tra  $H_{0j}$  e  $H_{1j}, j \neq i$ , sia vera quando si verifica l' $i$ -esima sotto-ipotesi.

L'assunzione (A.1) implica che l'insieme dei  $p$ -values  $\lambda_1, \dots, \lambda_k$  associato ai test parziali  $\mathbf{T}$  è *positivamente dipendente nell'alternativa*, irrispettivamente alle relazioni di dipendenza delle variabili casuali in  $\mathbf{X}$ .

## 6.5.2 PROPRIETÀ DESIDERABILI DELLE FUNZIONI DI COMBINAZIONE

La combinazione non parametrica in un test di secondo ordine  $T'' = \psi(\lambda_1, \dots, \lambda_k)$  è ottenuta con una funzione reale (misurabile) continua, monotona decrescente e non degenera  $\psi : (0,1)^k \rightarrow \mathfrak{R}^1$ . Per essere una buona funzione di combinazione,  $\psi$  deve soddisfare almeno le ragionevoli proprietà che seguono:

- (P.1) una funzione di combinazione  $\psi$  deve essere non crescente in ciascun argomento:  
 $\psi(\dots, \lambda_i, \dots) \geq \psi(\dots, \lambda'_i, \dots)$  se  $\lambda_i < \lambda'_i$ ,  $i \in \{1, \dots, k\}$ ;
- (P.2) ogni funzione di combinazione  $\psi$  deve raggiungere il suo estremo superiore  $\bar{\psi}$ , possibilmente non finito, quando almeno uno degli argomenti raggiunge lo zero:  
 $\lim_{\lambda_i \rightarrow 0} \psi(\dots, \lambda_i, \dots) = \bar{\psi}$ ,  $i \in \{1, \dots, k\}$ ;
- (P.3)  $\forall \alpha > 0$ , il valore critico di ogni  $\psi$  è assunto essere finito e strettamente inferiore al suo valore estremo superiore:  $T''_\alpha < \bar{\psi}$ ;

Queste proprietà definiscono una classe  $C$  di funzioni di combinazione; si può inoltre definire una sotto-classe  $C_S$  di funzioni di combinazione simmetriche<sup>11</sup>: appartengono a questa classe, ad esempio, le funzioni di combinazione di Fisher, di Liptak, di Tippett; presentiamo la loro espressione:

- $T''_F = -2 \cdot \sum_{i=1}^k \log(\lambda_i) \in C_S$ ; se i  $k$  test statistici parziali sono indipendenti e continui, allora sotto l'ipotesi nulla  $T''_F$  segue una distribuzione  $\chi^2$  centrale con  $2 \cdot k$  gradi di libertà;
- $T''_L = \sum_{i=1}^k \Phi^{-1}(1 - \lambda_i) \in C_S$ ; dove  $\Phi$  è la funzione di ripartizione della distribuzione normale standard; se i  $k$  test statistici parziali sono indipendenti e continui, allora sotto l'ipotesi nulla  $T''_L$  segue una distribuzione normale con media 0 e varianza  $k$ ;
- $T''_T = \max_{1 \leq i \leq k} (1 - \lambda_i) \in C_S$ ; se i  $k$  test statistici parziali sono indipendenti e continui,  $T''_T$  si comporta come il maggiore dei  $k$  valori casuali dalla distribuzione uniforme nell'intervallo aperto  $(0, 1)$ ;

E' utile ricordare che se i test parziali sono tutti test di permutazione esatti allora anche il test combinato  $T''_\psi$  è esatto  $\forall \psi \in C$ .

<sup>11</sup>  $C_S \subset C$  tali che se  $(u_1, \dots, u_k)$  è una qualunque permutazione di  $(1, \dots, k)$ , allora  $\psi(\lambda_{u_1}, \dots, \lambda_{u_k}) = \psi(\lambda_1, \dots, \lambda_k)$ ,  $\forall \psi \in C_S$ ;

## 6.5.3 ALGORITMO IN DUE FASI PER LA COMBINAZIONE NONPARAMETRICA

Si presenta ora un algoritmo in due fasi utilizzato per ottenere una stima Monte Carlo condizionata della distribuzione di permutazione dei test combinati. La prima fase riguarda la stima della distribuzione  $k$ -variata di  $\mathbf{T}$ , e la seconda trova la stima della distribuzione dei permutazione del test combinato  $T_{\psi}''$  utilizzando gli stessi risultati CMC ottenuti nella prima fase.

La prima fase dell'algoritmo include i seguenti passi:

(S.a<sub>k</sub>) si calcola il vettore dei valori osservati dei test  $\mathbf{T} : \mathbf{T}_o = \mathbf{T}(\mathbf{X})$ ;

(S.b<sub>k</sub>) si considera un elemento  $g^*$ , estratto casualmente da un gruppo appropriato di trasformazioni  $\mathbf{G}$ , e i valori delle statistiche vettore  $\mathbf{T}^* = \mathbf{T}(\mathbf{X}^*)$ , dove  $\mathbf{X}^* = g^*(\mathbf{X})$ . Nella maggior parte delle situazioni, la permutazione dei dati  $\mathbf{X}^*$  può essere ottenuta considerando prima una permutazione casuale  $(u_1^*, \dots, u_n^*)$  degli interi  $(1, \dots, n)$  e quindi assegnando i vettori dei dati individuali al gruppo appropriato;

(S.c<sub>k</sub>) si eseguono  $B$  ripetizioni indipendenti del passo (S.b<sub>k</sub>). L'insieme dei risultati CMC

$\{\mathbf{T}_r^*, r = 1, \dots, B\}$  è quindi un campione casuale dalla distribuzione di permutazione  $k$ -variata del vettore delle statistiche test  $\mathbf{T}$ ;

(S.d<sub>k</sub>) la funzione di ripartizione empirica (EDF)  $\hat{F}_B(\mathbf{z} | \mathbf{X}) = (\frac{1}{2} + \sum_{r=1}^B \mathbf{I}(\mathbf{T}_r^* \leq \mathbf{z})) / (B+1)$ ,  $\forall \mathbf{z} \in \mathfrak{R}^k$ , fornisce una stima della corrispondente distribuzione di permutazione  $k$ -dimensionale  $F(\mathbf{z} | \mathbf{X})$  di  $\mathbf{T}$ . Inoltre  $\hat{L}_i(z | \mathbf{X}) = (\frac{1}{2} + \sum_{r=1}^B \mathbf{I}(T_{ir}^* \geq z)) / (B+1)$ ,  $i = 1, \dots, k$ , fornisce una stima  $\forall z \in \mathfrak{R}^1$ , delle funzioni di livello di significatività marginali di permutazione  $L_i(z | \mathbf{X}) = \Pr\{T_i^* \geq z | \mathbf{X}\}$ ; quindi  $\hat{L}_i(T_i^{oss} | \mathbf{X}) = \hat{\lambda}_i$  fornisce una stima del  $p$ -value marginale  $\lambda_i = \Pr\{T_i^* \geq T_i^{oss} | \mathbf{X}\}$  relativo al test  $T_i$ ;

La seconda fase dell'algoritmo include i seguenti passi:

(C.a) i  $k$   $p$ -values osservati vengono stimati dai dati  $\mathbf{X}$  con  $\hat{L}_i(T_i^{oss} | \mathbf{X}) = \hat{\lambda}_i$ , dove  $T_i^{oss} = T_i(\mathbf{X})$

$i = 1, \dots, k$ , rappresenta il valore osservato del test parziale  $i$ -esimo, ed  $\hat{L}_i$  è l' $i$ -esima funzione di stima del livello di significatività ottenuta dalla procedura di ricampionamento condizionato sull'insieme di dati  $\mathbf{X}$ , cioè  $\hat{L}_i(T_i^{oss} | \mathbf{X}) = (\frac{1}{2} + \sum_{r=1}^B \mathbf{I}(T_{ir}^* \geq T_i^{oss})) / (B+1)$ ,  $\forall i = 1, \dots, k$ , dove  $B$  è il numero di ricampionamenti ed  $\mathbf{I}(\cdot)$  è la funzione indicatrice che vale 1 se  $(\cdot)$  è vera e zero altrimenti;

(C.b) il valore osservato del test combinato di secondo ordine è valutato ancora attraverso la stessa simulazione condizionata ed è dato da  $T_o'' = \psi(\hat{\lambda}_1, \dots, \hat{\lambda}_k)$ ;

(C.c) l' $r$ -esimo valore combinato delle statistiche vettore viene quindi calcolato da  $T_r'' = \psi(\lambda_{1r}^*, \dots, \lambda_{kr}^*)$ , dove  $\lambda_{ir}^* = \hat{L}_i(T_{ir}^* | \mathbf{X})$ ,  $i = 1, \dots, k$ ,  $r = 1, \dots, B$ ;

(C.d) quindi il  $p$ -value del test combinato  $T''$  è stimato da  $\hat{\lambda}_\psi'' = \sum_{r=1}^B \mathbf{I}(T_r'' \geq T''^{oss}) / B$ ;

(C.e) se  $\hat{\lambda}_\psi'' \leq \alpha$  allora si rifiuta  $H_0$  al livello di significatività  $\alpha$ ;

Un'importantissima proprietà asintotica di  $\hat{F}_B^*(z)$  dice che se  $B \rightarrow \infty \Rightarrow \hat{F}_B^*(z) \xrightarrow{q.c.} F(z | \mathbf{X})$ ,  $\forall z \in \mathfrak{R}^1$ ; (Pesarin, 2001; Proposizione 4.1, 4.2.1) essa ci assicura che il CMC fornisce delle stime consistenti e non distorte sia delle vere distribuzioni di permutazione  $F_\psi(t | \mathbf{X}) = \Pr\{T'' \leq t | \mathbf{X}\}$ ,  $\forall t \in \mathfrak{R}^1$ , sia del vero  $p$ -value  $\lambda_\psi'' = \Pr\{T'' \geq T'' | \mathbf{X}\}$ . Infatti, quando  $B \rightarrow \infty$ , si ha che  $\hat{\lambda}_i \rightarrow \lambda_i$  con probabilità 1,  $i = 1, \dots, k$ . Perciò anche  $\hat{\lambda}_\psi''$  converge a  $\lambda_\psi''$  con probabilità 1, dato che  $k$  è un intero fissato e la funzione di combinazione  $\psi$  viene assunta continua.

*Osservazione 6.7:* Con riferimento ai passi (S.d<sub>k</sub>), (C.a), (C.c) e (C.d), quando specificato, verranno utilizzate negli studi di simulazione dei capitoli 7 ed 8 delle stime dei  $p$ -values in accordo alla seguente definizione di livello di significatività osservato:

$$p_{obs} = \Pr\{T > T_{obs}; H_0\} + \frac{1}{2} \Pr\{T = T_{obs}; H_0\}, \quad \text{quindi, in generale,}$$

$$\hat{\lambda} = \hat{L}(z | \mathbf{X}) = \left\{ \sum_{r=1}^B \mathbf{I}(T^* > z + \frac{1}{2}) + \sum_{r=1}^B \mathbf{I}(T^* = z) \right\} / B.$$

#### 6.5.4 NON DISTORSIONE DEL TEST COMBINATO

In modo da ottenere la non-distorsione dei test combinati  $T_\psi'' \forall \psi \in C$ , si consideri prima il seguente lemma:

**Lemma 6.2, 6.3.2 (Pesarin, 2001):** Siano  $Y$  e  $Z$  due variabili aleatorie definite nello stesso spazio di probabilità univariato. Se  $Y$  è stocasticamente più grande di  $Z$ , così che le loro

funzioni di ripartizione soddisfano a  $F_Y(t) \leq F_Z(t)$ ,  $\forall t \in \mathfrak{R}^1$ , e se  $\varphi$  è una funzione reale (misurabile) non decrescente, allora  $\varphi(Y)$  è stocasticamente più grande di  $\varphi(Z)$ .

La dimostrazione di questo Lemma si basa sulla seguente relazione:

$$\Pr\{\varphi(Y) \leq \varphi(t)\} = F_Y(t) \leq F_Z(t) = \Pr\{\varphi(Z) \leq \varphi(t)\}, \quad \forall t \in \mathfrak{R}^1.$$

**Teorema 6.3, 6.3.2 (Pesarin, 2001):** Se, per un dato insieme di dati  $\mathbf{X}$  e un qualsiasi  $\alpha > 0$ , i test di permutazione parziali  $\mathbf{T} = \{T_i, i = 1, \dots, k\}$  sono marginalmente non distorti per  $H_{0i}$  contro  $H_{1i}$ ,  $i = 1, \dots, k$ , così che i loro  $p$ -values associati  $\lambda_i$ ,  $i = 1, \dots, k$ , sono positivamente dipendenti, allora  $T_\psi'' = \psi(\lambda_1, \dots, \lambda_k)$ ,  $\forall \psi \in C$ , è un test combinato non distorto per  $H_0 : \{\bigcap_{i=1}^k H_{0i}\}$  contro  $H_1 : \{\bigcap_{i=1}^k H_{1i}\}$ .

**Dimostrazione:** Le proprietà di non distorsione marginale e di dipendenza positiva dei test parziali  $T_i$  implicano che  $\Pr\{\lambda_i \leq z \mid \mathbf{X}(0)\} \leq \Pr\{\lambda_i \leq z \mid \mathbf{X}(\delta)\} \forall z \in (0, 1)$ ,  $i = 1, \dots, k$ , poiché i  $p$ -values  $\lambda_i(\mathbf{X}(\delta))$  sono stocasticamente più piccoli di  $\lambda_i(\mathbf{X}(0))$ . Quindi, per la proprietà (P.1) di non decrescenza della funzione di combinazione  $\psi$ , ed il Lemma 6.2, segue che  $\psi(\dots, \lambda_i(\mathbf{X}(0)), \dots)$  è stocasticamente più grande di  $\psi(\dots, \lambda_i(\mathbf{X}(\delta)), \dots)$ . Quindi, iterando da  $i = 1, \dots, k$ , si ottiene la non distorsione.

*Osservazione 6.8:* La non distorsione marginale dei test parziali è solo condizione sufficiente per la non distorsione dei test combinati.

*Osservazione 6.9:* Una importante conseguenza del Teorema 6.3, 6.3.2 è che se tutti i test parziali sono marginalmente non distorti  $\forall \alpha > 0$ , così che siano ordinati rispetto agli effetti trattamento, allora tutti i test nonparametrici combinati sono anche stocasticamente ordinati rispetto all'effetto trattamento.

### 6.5.5 CONSISTENZA DEL TEST COMBINATO

In modo da esaminare la consistenza del test combinato, si assuma inoltre che:

- (a) quando  $n \rightarrow \infty$ , allora anche le numerosità campionarie di tutti i gruppi  $n_j \rightarrow \infty$ ,  $j = 1, \dots, C$ ;
- (b) il numero  $B$  di iterazioni CMC diverge all'infinito;
- (c)  $k$  e  $\alpha$  sono fissati;

Si noti che quando il metodo CMC produce stime fortemente consistenti, il ruolo di  $B \rightarrow \infty$  risulta equivalente a considerare le vere quantità esatte di permutazione al posto delle loro stime. Inoltre si può assumere che la cardinalità  $M$  di  $\Omega_{/X}$  e il numero  $B$  di iterazioni CMC sono talmente elevate che tutti gli errori di approssimazione derivanti dalla sostituzione delle distribuzioni discrete con la loro immagine continua sono sufficientemente piccoli da essere trascurabili.

La consistenza del test combinato assumendo che (A.1) e (A.2) siano soddisfatte viene presentata in Pesarin (2001), Teorema 6.1, pagina 166. Verrà ora presentato un rilassamento di tale teorema.

Si assuma che (A.1) sia soddisfatto, e si rilassi (A.2) con il seguente:

(Ar.2) Almeno un test parziale di permutazione  $T_i$  è fortemente consistente per  $H_{0i}$  contro  $H_{1i}$ , cioè  $\Pr\{T_i \geq T_{i\alpha}\} \rightarrow 1, \forall \alpha > 0$ , quando  $n$  tende all'infinito, dove  $T_{i\alpha}$ , che è assunto essere finito, rappresenta il valore critico di  $T_i$  al livello  $\alpha$ ;

**Rilassamento del Teorema 6.1, 6.3.1 (Pesarin, 2001):** Se almeno un test parziale di permutazione  $T_i$  è fortemente consistente per  $H_{0i}$  contro  $H_{1i}$ , e la rispettiva sotto-ipotesi alternativa  $H_{1i}$  risulta vera,  $i \in \{1, \dots, k\}$ , allora  $T_{\psi}'' = \psi(\lambda_1, \dots, \lambda_k), \forall \psi \in C$ , è un test combinato fortemente consistente per  $H_0 : \{\bigcap_{i=1}^k H_{0i}\}$  contro  $H_1 : \{\bigcap_{i=1}^k H_{1i}\}$ .

**Dimostrazione:** Per essere fortemente consistente, il test combinato deve raggiungere la propria regione critica con probabilità 1 se almeno una sotto-ipotesi alternativa  $H_{1i}$  risulta vera,  $i \in \{1, \dots, k\}$ . Dalla ipotizzata consistenza forte del test parziale di permutazione  $T_i$  (Ar.2), e dal fatto che almeno  $H_{1i}$  risulta vera, allora  $\lambda_i \rightarrow 0$  con probabilità uno poiché  $n \rightarrow \infty$ . Quindi, per la Proposizione 4.1, 4.2.1 e le proprietà (P.2) e (P.3) segue che  $T_{\psi}'' \rightarrow \bar{\psi} > T_{\psi\alpha}''$ .

*Osservazione 6.10:* Si è richiesto che la sotto-ipotesi alternativa  $H_{1i}$  sia vera se il corrispondente test parziale di permutazione  $T_i$  viene ipotizzato essere fortemente consistente per  $H_{0i}$  contro  $H_{1i}$  poiché, se  $H_{1i}$  risultasse falsa ( $H_{0i}$  vera) allora  $\lambda_i$  risulterebbe uniformemente distribuito su  $(0,1)$ .

Si noti che le stesse conclusioni possono essere ottenute con la nozione di consistenza debole, che comunque risulta implicata dalla consistenza forte.

## 6.6 SCELTA DELLE STATISTICHE TEST PARZIALI DI PERMUTAZIONE

Oltre alle assunzioni (A.1) e (Ar.2) richieste per garantire la non distorsione e la consistenza del test combinato, le statistiche test parziali dovrebbero godere di una buona potenza, e quindi utilizzare statistiche sensibili all'aspetto considerato.

Le condizioni per una selezione ottima di una statistica test di permutazione  $T$  sono riportate in Lehmann (1986). Queste condizioni sono abbastanza restrittive e si realizzano molto raramente nel contesto nonparametrico. Essenzialmente, sono legate al fatto che una statistica test “migliore” debba essere espressa come funzione della distribuzione di popolazione  $P$ , ma nel contesto della verifica nonparametrica, questa distribuzione viene assunta non nota. Pertanto, una statistica test “migliore” per tutti gli insiemi dei dati  $\mathbf{X}$  e per tutte le distribuzioni di popolazione  $P$  generalmente non esiste, poiché non può essere univocamente determinata. Quindi, generalmente, i test di permutazione sono definiti in maniera euristica sulle basi di analisi intuitive. Comunque, molto spesso è possibile considerare la versione di permutazione delle statistiche basate sulla soluzione parametrica, tenendo in considerazione la seguente Proposizione, che regge sotto deboli condizioni di regolarità:

**Proposizione 3.1 (Pesarin, 2001):** Se  $T$  è la miglior statistica per una certa famiglia parametrica di distribuzioni e se la sua regione critica non dipende da una specifica alternativa, allora la sua controparte di permutazione è asintoticamente equivalente ad essa; quindi, poiché condividono la stessa funzione di potenza limite, la versione di permutazione di  $T$  risulta asintoticamente la migliore per la stessa famiglia.

Una dimostrazione formale di questa affermazione può essere trovata in Hoeffding (1952). Si può dunque affermare che, dato un test asintoticamente ottimo rispetto a un determinato principio, il test di permutazione basato sulla stessa statistica ne “eredita” le proprietà ottimali. Pertanto, per numerosità campionarie finite, la scelta del test di permutazione  $T$  basata su tali argomentazioni può essere considerata una *scelta ragionevole*.



*Osservazione 6.11:* Selezionare una statistica test basata sui dati è equivalente a lanciare una freccetta su un telo bianco e quindi disegnare il centro del bersaglio attorno al punto colpito, ed è categoricamente non valido (Berger, 2000).

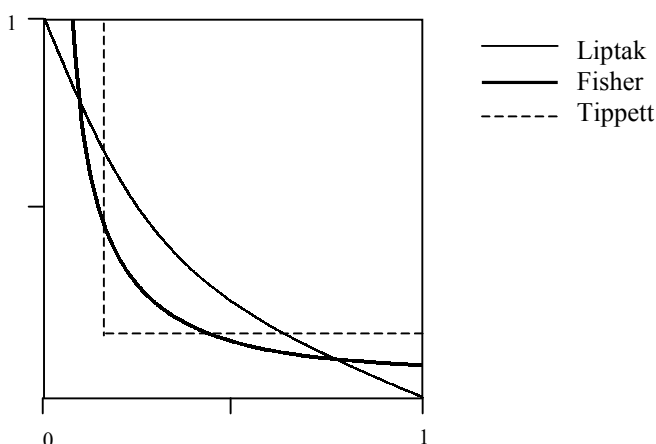
## 6.7 SCELTA DELLE FUNZIONI DI COMBINAZIONE

Nel contesto della metodologia di combinazione nonparametrica si presenta un problema: come trovare, per qualsiasi problema di verifica dato, la migliore funzione di combinazione in  $C$ . Ciò sembra molto difficile da stabilire e si crede sia irrisolvibile nel caso di numerosità campionarie finite e senza alcuna ulteriore restrizione.

Come accennato nel paragrafo 6.2 del presente capitolo, si può parlare di deviazioni “pure” quando solo uno (o pochi) degli aspetti considerati presenta una deviazione notevolmente significativa dall’ipotesi nulla, mentre di deviazioni “miste” quando molti o tutti gli aspetti considerati presentano deboli tendenze di significatività, ma che congiuntamente risultano informativi nel rendere l’ipotesi nulla meno probabile.

Pertanto, se a priori non sappiamo quanti dei  $k$  aspetti considerati presenteranno deviazioni dall’ipotesi nulla, la scelta di una funzione di combinazione  $\psi$  in  $C$  può sembrare un atto piuttosto arbitrario, poiché differenti funzioni di combinazione, a causa delle differenti regioni di rifiuto (figura 6.2) possono fornire  $p$ -values globali differenti, sebbene siano consistenti, e quindi asintoticamente equivalenti nell’alternativa.

**Figura 6.2:** Regioni critiche dei tre test combinati



La figura 6.2 descrive le regioni critiche delle funzioni di combinazione Fisher, Liptak e Tippett nella situazione molto semplice dove  $k = 2$  e i due test parziali sono indipendenti (Pesarin, 2001)

Tuttavia, le seguenti direttive pratiche descritte in Pesarin (2001) possono essere utili:

- Quando prevediamo che soltanto una o alcune, ma non tutte, le sotto-alternative possono risultare vere suggeriamo di usare la funzione di combinazione Tippett;
- Quando prevediamo che tutte le sotto-alternative possano essere congiuntamente vere, l'uso delle funzioni di combinazione Liptak o Diretta è generalmente giustificato;
- L'uso della funzione di combinazione Fisher è appropriato perché è generalmente intermedio tra quello della funzione Tippett e quello della funzione Liptak;



- Quando si preferisce una forma di combinazione neutra<sup>12</sup>, si può iterare la procedura di combinazione con funzioni differenti, affinché il  $p$ -value finale diventi ragionevolmente invariante;

## 6.8 ESEMPI IN LETTERATURA CONFORMI ALL'APPROCCIO MULTI-ASPECT

Si presenteranno quattro esempi trovati in letteratura, l'impostazione dei quali sottintende un approccio conforme alla logica multi-aspect.

ESEMPIO 6.1: ANALISI STATISTICA DI CONFIGURAZIONI SPAZIALI DI PUNTI CON COMBINAZIONE DI  $K$  INDICATORI (PESARIN, 1989)

L'ipotesi di casualità spaziale completa (*complete spatial randomness*, CSR) per una configurazione spaziale di punti afferma che :

- (i) il numero degli eventi in una qualsiasi regione piana  $A$  (finita) con area  $|A|$  segue una distribuzione di Poisson con media  $\lambda \cdot |A|$ ;
- (ii) dati  $n$  eventi  $\mathbf{x}_i$  in una regione  $A$ , gli  $\mathbf{x}_i$  sono un campione casuale indipendente da una distribuzione uniforme su  $A$ ;

Nella (i), la costante  $\lambda$  rappresenta l'intensità, o numero medio di eventi per area unitaria; la casualità spaziale completa implica quindi che l'intensità degli eventi non varia sul piano. In conformità con la (ii), la casualità spaziale completa afferma inoltre che non vi sono interazioni fra gli eventi, ovvero che l'esistenza di un evento  $E$  non incoraggia o inibisce l'avvenimento di altri eventi in prossimità di  $E$ .

Molte analisi iniziano con la verifica della casualità spaziale completa, per tre ragioni principali:

- (a) rifiutare la casualità spaziale completa è un prerequisito minimo per tentare di modellare una configurazione osservata;
- (b) i test vengono utilizzati per esplorare l'insieme dei dati e per aiutare nella formulazione di alternative plausibili alla casualità spaziale completa;
- (c) la casualità spaziale completa agisce come un'ipotesi che divide le configurazioni regolari da quelle aggregate;

Inoltre, in Diggle (1983), viene suggerito: (d) l'uso dei livelli di significatività ottenuti al fine di misurare la forza dell'evidenza contro la casualità spaziale completa; (e) la

---

<sup>12</sup> per la procedura iterata si veda il paragrafo 6.4;

combinazione informale di diversi test complementari, per indicare la natura di una qualsiasi deviazione dalla casualità spaziale completa.

Con riferimento al punto (e), se viene richiesta una singola valutazione di significatività, si può utilizzare la procedura “*maximum*” descritta nel paragrafo 6.2. Si supponga che i livelli di significatività raggiunti da  $k$  test non necessariamente indipendenti siano  $p_j = 1, \dots, k$  e sia  $p_{\min}$  il più piccolo tra essi, corrispondente alla deviazione maggiormente significativa dalla casualità spaziale completa. Quindi, sotto CSR:

$$p \leq \Pr\{p_{\min} \leq p\} \leq kp$$

Per  $k$  test indipendenti, risulta  $\Pr\{p_{\min} \leq p\} = 1 - (1 - p)^k$ .

Cox (1977) mette in evidenza che l'utilizzo del “*maximum*” test come parte di una procedura diagnostica ha senso pratico solo se i vari test utilizzati esaminano differenti aspetti della configurazione, così che un risultato significativo per un test non impedisce un'interpretazione sensata per gli altri.

Si è riscontrato (Diggle, 1983) che i metodi basati sulle *nearest neighbour distances* siano più potenti di quelli basati sulle *point to nearest event distances* per alternative regolari al CSR, e vice versa per alternative aggregate. Inoltre, un test basato su  $T_1$ , la distanza minima fra eventi<sup>13</sup>, è teoricamente preferibile per verificare il CSR contro certe alternative regolari, ma è potenzialmente sensibile alle imprecisioni nella registrazione dei dati. Un test basato su  $T_k$ , la  $k$ -esima distanza minima fra eventi<sup>14</sup>, soffre meno di questo difetto se  $k$  è maggiore di uno, ma coinvolge un elemento di arbitrarietà nella scelta di  $k$ . La statistica per i conteggi quadrati  $X^2$ <sup>15</sup> richiede una scelta più o meno arbitraria della dimensione e della forma dei *quadrats*, e fornisce un test che risulta debole contro alternative regolari.

L'analisi statistica di configurazioni spaziali di punti, da un punto di vista della verifica d'ipotesi, è basata sull'assunto che la configurazione osservata sia stata generata da un processo casuale sottostante. Si consideri la configurazione osservata come una realizzazione campionaria di un processo spaziale casuale. Si osservi che le procedure di verifica, per molti aspetti rilevanti, possono essere usate propriamente solo se le distribuzioni campionarie delle statistiche utilizzate siano note almeno approssimativamente. Quando il meccanismo casuale del processo spaziale di punti viene specificato completamente, sia per il modello probabilistico che per i suoi parametri, è possibile calcolare la distribuzione campionaria per ogni statistica di interesse. Comunque, come sottolineato da vari autori (Ripley, 1981; Diggle, 1983) il calcolo diretto di queste

<sup>13</sup> *minimum inter-event distance*;

<sup>14</sup> *k-th smallest inter-event distance* ;

<sup>15</sup> *quadrat count statistic*;

distribuzioni è generalmente poco pratico o estremamente difficile, ad eccezione del caso di casualità spaziale completa, cioè per il processo spaziale di Poisson omogeneo. D'altra parte, si possono eseguire esperimenti Monte Carlo per ottenere approssimazioni in tutti i casi in cui vi è completa specificazione del meccanismo casuale e di tutti i suoi parametri. Per i processi di Poisson non omogenei, come per altri processi, normalmente è impossibile ottenere la funzione di verosimiglianza a causa della sua struttura estremamente complicata.

Questo ha due principali conseguenze:

- (i) nessun criterio ottimo, nel significato rigorosamente decisionale, può essere propriamente seguito;
- (ii) l'analisi statistica si basa su un insieme di  $k$  indicatori diversi, che usualmente sono stabiliti con argomentazioni intuitive, qualitative o euristiche;

In Pesarin (1989) viene illustrato un metodo di simulazione nonparametrico, asintoticamente *model-free*, condizionato ad un'appropriata stima consistente della funzione di intensità semplice del processo sottostante. Un tale metodo è appropriato per l'analisi statistica di tutti i processi caratterizzati da una funzione di intensità semplice, cioè per il processo di Poisson non omogeneo e per tutti gli altri processi senza memoria. Il modello probabilistico ed i parametri (di disturbo) sono considerati come non noti.

Nell'analisi statistica di configurazioni spaziali di punti, normalmente sono prese in considerazione  $k$  diverse statistiche (indicatori), ciascuna informativa per un aspetto rilevante dell'analisi, e quindi, utile per il medesimo; tuttavia per problemi di verifica più generali, può accadere che tutte insieme siano congiuntamente informative, e pertanto si necessita una loro combinazione statistica.

Questa combinazione può essere facilmente ottenuta nel caso di indipendenza stocastica fra di esse, ma spesso essa non può essere invocata, e pertanto la combinazione deve essere nonparametrica, in particolare rispetto alla struttura non nota di dipendenza sottostante, nel senso che la distribuzione campionaria della statistica di secondo livello si ottiene irrispettivamente alla dipendenza sottostante fra le statistiche di primo livello.

Come esempio, si consideri il seguente insieme di dati spaziali espressi in conteggi quadrati (*quadrat counts*) relativi a due diversi tipi di oggetti osservati nella stessa regione finita (figura 6.3):

11 4	4 5	2 7
9 4	3 4	3 7

**Figura 6.3:**  
insieme di dati spaziali espressi in conteggi quadrati relativi a due diversi tipi di oggetti osservati nella stessa regione finita (Pesarin, 1989).

dove, per ogni quadrato,  $n_i$  ( $i = 1, \dots, 6$ ) indica il numero di oggetti del primo tipo ed  $m_i$  ( $i = 1, \dots, 6$ ) indica il numero di oggetti del secondo tipo.

A prima vista sembra che:

- a) il processo sottostante non è un processo omogeneo di Poisson;
- b) i dati mostrano piccole differenze nella concentrazione spaziale e forse nella somiglianza della distribuzione;

Quindi si possono analizzare separatamente entrambi gli aspetti, rispettivamente con un indice di differenza di concentrazione spaziale e con un indice di somiglianza di distribuzione. In questo esempio si potrebbe facilmente considerare anche statistiche basate sulle distanze fra gli eventi, se le loro coordinate fossero disponibili e se questo tipo di analisi risultasse adeguata.

indice di differenza di concentrazione:  $C = \left| \sum_{i=1}^6 n_i^2 / n^2 - \sum_{i=1}^6 m_i^2 / m^2 \right|$

indice di somiglianza di distribuzione: il Chi-quadrato convenzionale  $2 \times 6$

Si noti che entrambi gli indicatori sono stocasticamente crescenti al crescere delle discrepanze nel processo spaziale sottostante.

L'ipotesi nulla generale afferma che i due processi sottostanti sono spazialmente uguali eccetto per l'intensità totale, nel senso che la probabilità di assegnare un nuovo punto in qualsiasi regione è la stessa per i due processi, anche se spazialmente non costante.

Nell'analisi la relativa funzione di intensità è stimata da:

$$\tilde{L} = \{q_i = (n_i + m_i) / (nm), i = 1, \dots, 6\}$$

$n = 32$  e  $m = 31$  ( $n$  ed  $m$  sono considerate statistiche ancillari per il problema). Questa fornisce una stima, sotto l'ipotesi nulla, dei parametri medi del processo, assumendo che ogni sotto-regione sia approssimativamente CSR e che i sei sotto-processi siano indipendenti.

Fissando  $B = 1000$  otteniamo:  $C = 0.05551$ ,  $\chi^2 = 9.80809$ ;  $\lambda_1 = 0.063$ ,  $\lambda_2 = 0.07$ ;

$U = 10.76577$ ,  $\lambda = 0.045$ ;

Si osservi che per verificare una generale differenza spaziale dei due processi sottostanti, entrambi gli indicatori risultano marginalmente non significativi per  $\alpha = 0.05$ , mentre il test combinato risulta significativo.

ESEMPIO 6.2: CASUALITÀ E TEST STATISTICI DI CASUALITÀ NELL'ANALISI CRITTOGRAFICA  
(PESARIN, 1989)

Tutti i test statistici per la verifica di un'insieme di ipotesi sono basati su una sorta di comune principio: tutti cercano di mettere a confronto quanto tramite indicatori viene osservato direttamente sui dati con quanto ci si dovrebbe attendere sulla base dell'aspettativa conforme alla ipotesi saggiata (di casualità, nel caso in esame).

Si consideri un modello di completa casualità: il modello stocastico di assegnazione di probabilità ad un messaggio  $M$  <sup>(16)</sup> suppone che *ogni carattere della successione venga estratto a caso dalla distribuzione di probabilità  $p_L(a_i), i = 1, \dots, m$  dei caratteri alfabetici della lingua  $L$ .*

La distribuzione  $p_L(a_i)$  viene stimata con la frequenza relativa  $\hat{p}_L = f_M(a_i)/n$  <sup>(17)</sup>,  $1 \leq i \leq m$ , da un testo campione  $M$  di lunghezza  $n$  sufficientemente grande. Secondo questo modello, il messaggio  $M$  viene assimilato alla realizzazione di un processo stocastico bernoulliano a componenti indipendenti e ugualmente distribuite (puro rumore bianco, processo puramente casuale) in cui si immagina che i successivi caratteri vengano estratti a caso con reinserimento da un'urna contenente palline contrassegnate in proporzione a  $p_L(a_i)$ . Per cui:

$$P_L(M) = P_L(x_1, \dots, x_n) = \prod_{j=1}^n [p_L(a_i)]^{f_M(a_i)}$$

Si tratta di un modello alquanto semplicistico che tuttavia, pur se inadeguato ad una buona descrizione stocastica dei messaggi, consente di costruire strumenti statistici utili nella decrittazione di cifrati con sistemi di sostituzione polialfabetica a periodicità finita (Konheim, 1978). In sostanza,  $P_L(M)$  fornisce la cosiddetta funzione di verosimiglianza del messaggio  $M$  nella lingua  $L$ , dato il modello bernoulliano.

In generale, i messaggi cifrati  $C = E(M; K)$  sono ottenuti con metodi di sostituzione ordinata, ossia con algoritmi tali per cui il processo del cifrato è la combinazione del processo chiaro e di un processo ausiliario di occultamento, dipendente da una chiave  $K$ , secondo una forma del tipo:

$$c_i = x_i + \varepsilon_i(K) [\text{mod } m], \quad 1 \leq i \leq n$$

<sup>16</sup> Un messaggio  $M \in \mathcal{M}$ , ( $\mathcal{M}$  è l'insieme dei possibili messaggi), lungo  $n$ , in un alfabeto  $L$  viene interpretato come una successione di  $n$  simboli o caratteri dell'alfabeto:  $M = \{x_1, \dots, x_n\}$  in  $L = \{a_1, \dots, a_m\}$ , ove  $m$  è il numero di simboli di  $L$ ;

<sup>17</sup> Dato un testo campione di lunghezza  $n$  sufficientemente grande, la distribuzione di frequenza dei caratteri di  $L$  nel testo campione  $M$  è:  $f_M(a_i) = \sum_{1 \leq j \leq n} \mathbf{I}(x_j; a_i)$ ,  $1 \leq i \leq m$  dove  $\mathbf{I}$  rappresenta la funzione contatore:  $\mathbf{I}(x_j; a_i) = 1$  se  $x_j = a_i$  e 0 altrimenti;

in cui per convenzione  $a_0 = a_m$  e  $0 = m$ . Per tali sistemi di cifratura, se il processo  $\varepsilon$  non ha periodicità bassa (meglio se è aperiodico e meglio ancora se è aleatorio) la distribuzione dei caratteri si presenta in genere più appiattita nel cifrato rispetto a quella del chiaro:  $f_C$  è meno concentrata di  $f_M$ . Anzi, se  $\varepsilon$  è un processo aleatorio a componenti indipendenti ed uniformemente distribuite (Rizzi, 1987), ossia è un processo puramente casuale, qualunque siano il messaggio  $M$ , il processo che lo ha generato e la lingua in cui è scritto. Questo risultato supporta bene l'aspettativa per cui il metodo di cifratura statisticamente ideale è il risultato della somma, modulo  $m$ , del messaggio con un processo di occultamento (pseudo) casuale a componenti stocasticamente indipendenti ed uniformemente distribuite.

In quanto segue, si indicherà con  $Y = \{y_1, \dots, y_n\}$  la sequenza di  $n$  caratteri da esaminare statisticamente sotto il profilo della casualità, per cui  $Y$  è sia  $M$ , sia  $C$  e sia  $\varepsilon$ ; gli elementi  $y$  di  $Y$  potranno essere indifferentemente trattati come simboli o come valori numerici nella corrispondenza  $\{a_i\} = \{i\}$ ,  $i = 1, \dots, m$ ;  $f_i = f_Y(a_i)$  indicherà la frequenza del carattere  $a_i$  in  $Y$ ; ecc.

Test noti sono, ad esempio, il test sulle frequenze dei caratteri, il test sulle frequenze di transizione, il test sulle distribuzioni in sottosequenze ed il test di dipendenza seriale modulo  $m$ . (per una rassegna dei principali test di casualità disponibili, si veda Baldessari, 1987 e Rizzi, 1987).

Si supponga, ora, di aver calcolato sulla sequenza osservata  $Y$ , la batteria di  $k$  test  $T_1, \dots, T_k$ , ciascuno a suo modo specializzato nel porre in luce un proprio aspetto della casualità. Sorge quindi l'esigenza di combinarne l'apporto informativo in una sorta di test di secondo grado. Essendo calcolati tutti sulla medesima sequenza osservata, in generale essi sono tra loro stocasticamente dipendenti, in quanto funzione dei medesimi dati, a meno che non si riesca a provarne positivamente l'indipendenza. Per di più, in generale la forma di dipendenza è spesso troppo complessa da potersi studiare e gestire in via analitica, e quindi, deve essere considerata come non nota. Di qui la necessità di procedere ad una loro combinazione in forma non parametrica rispetto alla sottostante struttura di dipendenza non nota.

#### ESEMPIO 6.3: TEST DI ADATTAMENTO PER DISTRIBUZIONI DI CUI SIANO NOTE FREQUENZE E QUANTITÀ (DANCELLI, 1991)

Le statistiche di tipo Chi-Quadrato sono tra i test più impiegati per valutare la bontà di adattamento di modelli teorici a distribuzioni empiriche. Tali statistiche operano un confronto tra frequenze osservate e attese nelle classi; tale modo di procedere può però apparire riduttivo quando delle classi siano note non solo le frequenze ma anche le

quantità, come avviene ad esempio nel caso delle distribuzioni dei redditi. Nella convinzione che questo secondo importante elemento di informazione non vada trascurato, Dancelli (1991) considera una nuova statistica,  $V^2$ , la quale propone un confronto fra medie osservate e attese nelle classi.

Si supponga senza perdita di generalità, che la caratteristica di interesse presente sulla popolazione sia interpretata dalla variabile casuale continua  $X$ , con funzione di ripartizione  $F(x)$ . Nel contesto in esame si assume che l'ipotesi da verificare sia  $H_0 : F(x) = F_0(x)$  dove  $F_0(x)$  è completamente specificata. Si assume inoltre che l'intervallo di variazione di  $X$  sia suddiviso in  $g$  classi contigue  $[x_{i-1}, x_i)$ , dove gli estremi sono prefissati. Essendo per ipotesi  $F_0(x)$  completamente specificata, è possibile calcolare i seguenti valori teorici:

$$\pi_i = \int_{x_{i-1}}^{x_i} dF$$

$$\mu_i = (1/\pi_i) \int_{x_{i-1}}^{x_i} x dF$$

$$\sigma_i^2 = (1/\pi_i) \int_{x_{i-1}}^{x_i} (x - \mu_i)^2 dF$$

dove  $i = 1, \dots, g$ .

Sia  $(X_1, \dots, X_n)$  un campione bernoulliano di  $n$  elementi scelto su  $X$ . Su di esso si definiscono le seguenti variabili casuali:

$$N_i = \# [X_j \in [x_{i-1}, x_i)]$$

$$P_i = N_i / n$$

$$X_i = (1/N_i) \sum_{w_i} Y_j \quad \text{dove } w_i = \{j : X_j \in [x_{i-1}, x_i)\}; i = 1, \dots, g.$$

$$V^2 = \sum_{i=1}^g \left( \frac{\bar{X}_i - \mu_i}{\sigma_i / \sqrt{n\pi_i}} \right)^2$$

Dancelli precisa che  $V^2$  non vuole necessariamente rappresentare un'alternativa ai tradizionali test basati sulle frequenze. L'idea di base è quella di utilizzare informazioni supplementari, non sostitutive, e si pensa alla possibilità di far ricorso a test che impiegano congiuntamente statistiche dell'uno e dell'altro tipo. Viene mostrato che  $V^2$  presenta una distribuzione limite di una variabile casuale Chi-quadrato con un numero di gradi di libertà pari al numero delle classi in cui è suddivisa la distribuzione. Per quanto riguarda l'impiego congiunto di statistiche costruite su quantità e frequenze, esso può

evidentemente realizzarsi in molti modi. Ad esempio, con riferimento alle due statistiche considerate in questo lavoro, un criterio estremamente semplice potrebbe condurre alla costruzione della variabile casuale  $S^2 = V^2 + X^2$  il cui comportamento asintotico in caso di ipotesi semplice può essere dedotto sin d'ora. Avendo infatti dimostrato l'indipendenza asintotica di  $V^2$  e  $X^2$ , per la proprietà riproduttiva della variabile casuale Chi-quadrato,  $S^2$  presenta sotto  $H_0$  la distribuzione asintotica di una variabile casuale Chi-quadrato centrale con  $2g - 1$  gradi di libertà e sotto  $H_1$  quella di una variabile casuale Chi-quadrato non centrale con  $2g - 1$  gradi di libertà e parametro di non centralità  $\lambda + \delta$ .

ESEMPIO 6.4: TEST NON PARAMETRICI BI-ASPECT PER PROBLEMI DI POSIZIONE CON APPLICAZIONE A STATISTICHE MEDICHE, DI BUSINESS E INDUSTRIALI (MAROZZI, 2002)

Partendo dalla considerazione che per i problemi di posizione a due campioni sia il test  $t$  di Student sia il test di permutazione tradizionale ( $T_a = \sum_{i=1}^{n_2} X_{2i}$ ) possono essere visti come test *uni-aspect*, poichè considerano solo l'aspetto "numerico"  $X_{ji}$  di ogni unità campionaria, Marozzi (2002) propone il test di permutazione *bi-aspect*  $T_{ab}$ , ideato come combinazione del test di permutazione tradizionale  $T_a$  con un test semplice che considera se un'osservazione campionaria sia "grande" o "piccola".

Il modello considerato per il problema in questione è un modello di risposta omoschedastico con effetti addittivi fissi<sup>18</sup>; si assume perciò che sotto  $H_1$  la distribuzione di  $X_2$  sia spostata di una quantità  $\delta > 0$  rispetto a quella di  $X_1$ , così che  $H_1 : \{X_1 \stackrel{d}{=} (X_2 - \delta)\}$ . Con riferimento ad alternative unidirezionali, le ipotesi possono essere scritte in maniera equivalente come  $H_0 : \{\delta = 0\}$  e  $H_1 : \{\delta > 0\}$ .

Il test  $T_a$  considera solo l'aspetto quantitativo  $X_{2i}$  di ogni elemento del secondo campione, quello che Marozzi chiama l'aspetto "numerico"; per costruire test  $T_{ab}$ , invece, viene considerato un altro aspetto quantitativo delle unità campionarie: questo secondo aspetto è relativo alla classificazione delle unità campionarie in due gruppi, uno con unità "grandi" e l'altro con unità "piccole". Un modo per formalizzare questa classificazione consiste nel definire "grande" un'osservazione maggiore della mediana  $\tilde{M}$  di  $\mathbf{X} = \{X_{ji}, i = 1, \dots, n_j; j = 1, 2\}$  e "piccola" altrimenti. Marozzi chiama questo secondo aspetto "categoriale".  $\tilde{M}$  viene definita come:

<sup>18</sup> per una descrizione formale di tale modello si veda il capitolo 8, paragrafo 8.2.1;



$$\tilde{M} = \begin{cases} X_{((n+1)/2)} & \text{se } n = n_1 + n_2 \text{ è dispari;} \\ 0.5 \cdot (X_{(n/2)} + X_{((n/2)+1)}) & \text{se } n = n_1 + n_2 \text{ è pari;} \end{cases}$$

dove  $X_{(i)}$  rappresenta l' $i$ -esima statistica ordinata del campione raggruppato. La tabella 6.5 rappresenta i risultati di questa classificazione:

**Tabella 6.5:** Classificazione delle unità campionarie nei due gruppi considerati (Marozzi, 2002).

	primo campione	secondo campione	
gruppo di unità “piccole”	$n_{1p}$	$n_{2p}$	$n_p$
gruppo di unità “grandi”	$n_{1g}$	$n_{2g}$	$n_g$
	$n_1$	$n_2$	$n$

Sotto  $H_0$ , ci si aspetta che le quantità  $n_{1g} / n_1$  e  $n_{2g} / n_2$  non siano molto differenti, mentre sotto  $H_1$ , ci si aspetta che al crescere di  $\delta$ , cresca la differenza tra le due quantità. Si noti che  $n_g$  (e quindi anche  $n_p$ ) risulta invariante alle permutazioni di  $\mathbf{X}$ . Una statistica test naturale e conveniente per studiare il secondo aspetto è

$$T_b = n_{2g} = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i} > \tilde{M}) \quad \text{dove } \mathbf{I}(\cdot) \text{ rappresenta la funzione indicatrice.}$$

In corrispondenza al modello ipotizzato l'ipotesi nulla globale può essere equivalentemente espressa come verifica dell'uguaglianza di due distribuzioni:

$H_0 : \{P_1^d = P_2^d\}$  dove  $P_j$  rappresenta la distribuzione della  $j$ -esima popolazione ( $j = 1, 2$ ); la decomposizione di  $H_0$  nelle due sotto ipotesi risulta:

$$H_{0a} : \{\delta = 0\} \text{ contro } H_{1a} : \{\delta > 0\} \text{ e}$$

$$H_{0b} : \Pr\{X_1 > \tilde{M}\} = \Pr\{X_2 > \tilde{M}\} \text{ contro } H_{1b} : \Pr\{X_1 > \tilde{M}\} > \Pr\{X_2 > \tilde{M}\}$$

Entrambe le statistiche  $T_a$  e  $T_b$  sono significative per valori elevati e forniscono informazioni sul problema di verifica: attraverso la combinazione nonparametrica sui test dipendenti, si può combinare le informazioni fornite dai due test in un test globale:

$$T_{ab} = \psi(\lambda_a, \lambda_b) = \max(1 - \lambda_a, 1 - \lambda_b)$$

dove  $\lambda_a$  e  $\lambda_b$  rappresentano i  $p$ -values corrispondenti rispettivamente ai due test parziali  $T_a$  e  $T_b$ , mentre  $\psi$  rappresenta la funzione di combinazione di Tippett. La scelta della funzione di combinazione viene motivata dal fatto che essa sotto  $H_1$  seleziona il miglior  $p$ -value ottenuto dai test parziali (vale a dire il  $p$ -value più piccolo), in conformità con la scelta di proporre un test con buone prestazioni generali, che possa quindi essere utilizzato senza la preoccupazione di quale sia la distribuzione sottostante ai dati, la quale di solito risulta non nota.

La finalità del test bi-aspect consiste infatti nel considerare un test che si comporti come l'usuale test di permutazione  $T_a$  quando il totale campionario rappresenta una statistica corretta per studiare la posizione, e che possa funzionare meglio di  $T_a$  altrimenti, in virtù dell'informazione aggiuntiva fornita dalla statistica test parziale  $T_b$ . Si tenga in considerazione che:

- 1) Il metodo il più comunemente usato per confrontare le posizioni di due popolazioni è il test  $t$  di Student. Quando le popolazioni sono normalmente distribuite e omoschedastiche, questo test è il test simile uniformemente più potente per verificare l'ipotesi nulla di posizioni uguali contro l'ipotesi alternativa unilaterale (Lehmann, 1986). E' inoltre esatto, non distorto e consistente. Se la distribuzione sottostante è differente, il test  $t$  di Student perde queste proprietà e le relative prestazioni possono essere scarse;
- 2) Per distribuzioni asimmetriche e/o con valor medio inesistente e/o in presenza di dati anomali, la mediana risulta essere una statistica più sensibile della media (utilizzata dai test tradizionali  $t$  e  $T_a$ ) nel rilevare spostamenti di posizione;
- 3) È noto (Johnson & Kotz, 1970) che con distribuzioni Cauchy o *half*-Cauchy, per le quali non esiste valor medio finito, la media del campione (che è permutazionalmente equivalente al totale del campione) non è una statistica adeguata per stimare la posizione. Inoltre, non esiste un test parametrico uniformemente più potente (UMP) per verificare  $H_0 : \{\delta = 0\}$  contro  $H_1 : \{\delta > 0\}$  per distribuzioni *half*-Cauchy o Cauchy (Lehmann, 1986);

Marozzi (2002) dimostra attraverso uno studio via simulazione che il test bi-aspect risulta notevolmente più potente del tradizionale test parametrico e di quello di permutazione per distribuzioni con “code pesanti” come la Cauchy, la *half*-Cauchy, e alcune miscele di distribuzioni normali con variabilità diseguali (vale a dire distribuzioni “contaminate” da una proporzione (10% e 30%) di dati anomali o *outliers*). Per la distribuzione esponenziale ed esponenziale doppia, il test proposto risulta migliore del test di permutazione tradizionale, mentre per la distribuzioni normale, uniforme, Chi-Quadrato e

una distribuzione bimodale, il test bi-aspect si comporta come il test di permutazione tradizionale (Tabella 6.6).

**Tabella 6.6:** Stima della potenza relativa di  $T_{ab}$  rispetto a  $T_a$  (Marozzi, 2002)

		$n_1$			$T_a$			
		10	20	40	10	10	20	40
		$n_2$			20	40	10	10
<i>distribuzioni</i>								
<i>simmetriche</i>	<i>Normale</i> $N(0, 1)$	0.98	0.99	0.98	0.96	0.93	0.95	0.92
	<i>Uniforme</i> $U(0, 2\sqrt{3})$	0.98	0.98	0.95	0.92	0.91	0.93	0.90
	<i>Esponenziale doppia</i> $L(0, 1/\sqrt{2})$	1.06	1.11	1.21	1.11	0.98	1.10	0.98
	<i>Cauchy</i> $(0, 1)$	1.54	1.97	1.90	1.78	1.68	1.79	1.71
	<i>Bimodale: mistura</i> $N(-1.5, 1)$ e $N(1.5, 1)$	0.98	0.98	0.94	0.94	0.91	0.93	0.90
<i>asimmetriche</i>	<i>Half-Cauchy</i> $(0, 1)$	1.40	1.86	2.12	1.99	1.98	1.32	1.09
	<i>Chi-quadrato n.c.</i> $\chi^2(0.25, 8)$	1.00	1.00	0.98	1.01	0.95	0.96	0.92
	<i>Esponenziale</i> $Exp(1)$	1.04	1.05	1.05	1.17	1.20	0.99	0.95
	<i>Contaminate (10%)</i> : mistura $N(0,1)$ e $N(1,100)$	1.30	1.70	1.89	1.60	1.53	1.60	1.51
	<i>Contaminate (30%)</i> : mistura $N(0,1)$ e $N(1,100)$	1.61	1.95	1.98	1.71	1.47	1.75	1.47

Nel capitolo 8, paragrafo 8.8, verrà considerato uno studio comparativo al test bi-aspect proposto da Marozzi per il problema della verifica dell'uguaglianza di due distribuzioni nel caso di un modello di risposta omoschedastico con effetti addittivi fissi.

# Capitolo 7

## Potenza statistica quando si esamina la differenziazione genetica: un'applicazione della procedura multi-aspect

### 7.1 INTRODUZIONE

L'evoluzione è il cambiamento genetico che si realizza nel corso del tempo in una popolazione di organismi. Essa si riferisce non a cambiamenti che avvengono in un individuo durante la sua vita, ma a cambiamenti delle caratteristiche delle popolazioni che si realizzano durante molte generazioni. Ciascuna popolazione possiede un *pool genico* isolato che è rappresentato dal materiale genetico di tutti gli individui appartenenti alla popolazione. Il *pool genico* comprende tutti i possibili alleli di ciascun gene della popolazione. Poiché molte specie sono diploidi, ogni individuo di una popolazione possiede solo due alleli per locus, uno per ciascun cromosoma omologo. Se una popolazione non è in evoluzione, le frequenze di ogni allele del *pool genico* restano costanti da una generazione a una generazione successiva. Tuttavia nel corso dell'evoluzione alcuni alleli aumentano di frequenza mentre altri diminuiscono o addirittura scompaiono. In altre parole le popolazioni evolvono attraverso variazioni delle frequenze alleliche, cioè per cambiamenti delle proporzioni relative dei loro vari alleli.

### 7.2 PRESENTAZIONE DEL PROBLEMA

Numerose tecniche statistiche vengono applicate in questo tipo di problema, ma non si è ancora sviluppata un'accurata discussione circa i loro relativi meriti nel rilevare l'eterogeneità genetica, mancanza particolarmente evidente per i metodi che *combinano* le informazioni di più loci.

In una situazione tipica il ricercatore ha raccolto dei campioni di tessuto da due o più gruppi di individui che sono separati nello spazio o nel tempo. L'applicazione di alcune tecniche biochimiche o molecolari fornisce i genotipi di uno o più loci nucleari o del genoma mitocondriale e ogni campione viene descritto in termini di frequenze alleliche.

Si supponga di avere una variabile risposta  $Y$  ed un insieme di osservazioni suddiviso in due gruppi, cosicché i dati possono essere rappresentati nel seguente modo:

$$\mathbf{Y} = \{Y_{hji}, i = 1, \dots, 2 \cdot n_j, j = 1, 2, h = 1, \dots, k\}$$

dove  $h$  indica il locus considerato,  $j$  il gruppo di appartenenza ed  $n_j$  il numero di soggetti in ciascun gruppo; la variabile  $Y_{hji}$  assume due modalità di risposta,  $a_h$  e  $A_h$ , in corrispondenza rispettivamente dell'allele comune e di quello raro che si possono presentare nell' $h$ -esimo locus.

Si tenga presente che ogni individuo considerato è diploide, e quindi possiede due alleli per locus (locus biallelico): a  $n_j$  ( $j = 1, 2$ ) individui considerati corrisponderanno  $2 \cdot n_j$  alleli.

*Osservazione 7.1:* In questo contesto si assume soddisfatto l'equilibrio di Hardy-Weinberg, pertanto i due alleli rilevati nel  $h$ -esimo locus del medesimo individuo vengono considerati come indipendenti. Dovrebbe essere osservato che, a causa della ricombinazione, del complesso processo di selezione o della bassa numerosità campionaria, l'equilibrio Hardy-Weinberg e, quindi, l'ipotesi di indipendenza può reggere soltanto approssimativamente. Tuttavia, benché la legge Hardy-Weinberg sia abbastanza restrittiva, « *in practice it is almost always found to apply to a quite satisfactory approximation* » ( Smith, 1983, p.681 )<sup>1</sup>.

Si consideri una variabile risposta binaria tale che:

$$X_{hji} = \mathbf{I}(Y_{hji} = A_h) = \begin{cases} 0 & \text{se } Y_{hji} = a_h \\ 1 & \text{se } Y_{hji} = A_h \end{cases} \quad i = 1, \dots, 2 \cdot n_j, j = 1, 2, h = 1, \dots, k$$

I due campioni verranno pertanto descritti in termini di frequenze alleliche per ogni locus considerato (tabella 7.1):

**Tabella 7.1:** Frequenza degli alleli  $\{A_h, a_h\}$  nell' $h$ -esimo locus

	$A_h$	$a_h$	<i>totale</i>
<i>1° campione</i>	$f_{A_h1}$	$f_{a_h1}$	$2 \cdot n_1$
<i>2° campione</i>	$f_{A_h2}$	$f_{a_h2}$	$2 \cdot n_2$
<i>totale</i>	$f_{A_h}$	$f_{a_h}$	$2 \cdot (n_1 + n_2)$

<sup>1</sup> Per una descrizione maggiormente accurata della legge di Hardy-Weinberg si veda il paragrafo 7.8;

dove  $f_{A_h,j} = \sum_{i=1}^{2 \cdot n_j} \mathbf{I}(Y_{hji} = A_h) = \sum_{i=1}^{2 \cdot n_j} X_{hji}$  ( $h = 1, \dots, k, j = 1, 2, i = 1, \dots, 2 \cdot n_j$ ) rappresenta la frequenza assoluta dell'allele  $A_h$  nell' $h$ -esimo locus del  $j$ -esimo campione.

Le specifiche domande scientifiche possono variare da studio a studio, ma una basilare, che determina come procedere con le analisi è la seguente: “Sono le differenze di frequenza degli alleli osservate fra i campioni abbastanza grandi da suggerire che essi non siano estratti dalla stessa popolazione (*pool genico*)?”

Sembra che nella maggior parte dei casi il modello evolutivo sottostante sia uno fra “*selective neutrality - isolation - genetic drift*”, i quali implicano che tutti i loci polimorfi esaminati siano potenzialmente informativi in riferimento alla domanda di eterogeneità genetica generale.

L'approccio statistico utilizzato più frequentemente consiste nell'effettuare in un primo passo, separatamente per ogni locus considerato, un test sulle tabelle di contingenza per verificare l'omogeneità di frequenza dell'allele, e in un secondo passo nel valutare simultaneamente, o congiuntamente, le informazioni di tutti i loci esaminati (tabella 7.2):

**Tabella 7.2:** approccio statistico in due passi

<i>locus</i>	<i>primo passo</i> <i>sotto-ipotesi</i>	<i>secondo passo</i> <i>ipotesi globale</i>
1	$H_{01} : \{X_{11} \stackrel{d}{=} X_{12}\}$ contro $H_{11} : \{X_{11} \stackrel{d}{\neq} X_{12}\}$	$H_0 : \{\bigcap_{h=1}^k H_{0h}\}$ contro $H_1 : \{\bigcup_{h=1}^k H_{1h}\}$
...	...	
$h$	$H_{0h} : \{X_{h1} \stackrel{d}{=} X_{h2}\}$ contro $H_{1h} : \{X_{h1} \stackrel{d}{\neq} X_{h2}\}$	
...	...	
$k$	$H_{0k} : \{X_{k1} \stackrel{d}{=} X_{k2}\}$ contro $H_{1k} : \{X_{k1} \stackrel{d}{\neq} X_{k2}\}$	

dove il simbolo ‘ $\neq$ ’ significa ‘ $<$ ’, ‘ $\neq$ ’ o ‘ $>$ ’.

*Osservazione 7.2:* La procedura di verifica sopra descritta può essere interpretata in conformità all'approccio multi-aspect “esteso”, nel quale ogni singola variabile (locus) rappresenta un aspetto del problema.

In letteratura uno studio che valuta la capacità di rilevare eterogeneità genetica combinando le informazioni di più loci si può trovare in Ryman e Jorde (2001), nel quale vengono confrontati due approcci statistici, il *summation method* contro l'applicazione del metodo di Bonferroni.

### 7.3 ESEMPLIFICAZIONE DEL PROBLEMA

Ryman e Jorde presentano (tabella 7.3) le frequenze alleliche di 12 loci allozimi biallelici codominanti per due consecutive coorti annuali di trota marrone (*Salmo trutta*) del lago Blanktjärnen, in Svezia centrale (per i particolari si veda Jorde & Ryman, 1996). Le differenze di frequenza degli alleli osservate sono abbastanza grandi da suggerire che ci siano vere differenze genetiche fra le due coorti annuali?

**Tabella 7.3:** Frequenze alleliche relative e statistiche test per tabelle di contingenza  $2 \times 2$  per 12 loci allozimi biallelici per due coorti (1992 e 1993) di trota marrone del lago Blanktjärnen, Svezia. Il numero di pesci è 43 e 27 nel 1992 e 1993, rispettivamente. (Ryman e Jorde, 2001)

locus	frequenza allele coorte		test esatto di Fisher		test Chi-quadrato		test Chi-quadrato correzione di Yates	
	1992	1993	$p$	$-2 \ln(p)$	$X^2$	$p$	$X_C^2$	$p$
<i>sAAT-4</i>	0.6395	0.4074	0.009	9.241	7.222	0.007	6.331	0.012
<i>DIA</i>	0.7791	0.8148	0.673	0.793	0.258	0.611	0.087	0.768
<i>bGALA-2</i>	0.7791	0.9444	0.009	9.489	6.849	0.009	6.298	0.012
<i>bGLUA</i>	0.8256	0.7407	0.285	2.512	1.454	0.228	0.969	0.325
<i>G3PDH-2</i>	0.7558	0.6296	0.129	4.097	2.550	0.110	1.954	0.162
<i>sIDHP-1</i>	0.4302	0.5741	0.119	4.261	2.748	0.097	2.207	0.137
<i>LDH-5</i>	0.7209	0.6852	0.705	0.700	0.205	0.651	0.068	0.794
<i>aMAN</i>	0.9884	0.9259	0.073	5.240	3.756	0.053	2.099	0.147
<i>sMDH-2</i>	0.9767	0.9815	1.000	0.000	0.036	0.851	0.165	0.685
<i>ME</i>	0.8953	0.8704	0.786	0.482	0.205	0.651	0.032	0.858
<i>MPI</i>	0.1047	0.2222	0.087	4.881	3.596	0.058	2.662	0.103
<i>PEPLT</i>	0.6395	0.6667	0.856	0.311	0.107	0.743	0.021	0.885
		<i>somma</i>		42.19	28.98		22.89	
		<i>somma dei g.d.l.</i>		24	12		12	
		$p$		0.012	0.004		0.029	

La differenza di frequenza dell'allele osservata tra i due campioni su ogni locus specifico (quindi la singola tabella di contingenza  $2 \times 2$ ) è stata esaminata utilizzando i seguenti test parziali:

- il test delle probabilità esatte di Fisher;
- il test Chi-quadrato di Pearson per tabelle di contingenza  $2 \times 2$  ( $X^2$ , con un grado di libertà, g.d.l. = 1);
- il test Chi-quadrato con la correzione di continuità di Yates ( $X_C^2$ , g.d.l. = 1);

Entrambe le approssimazioni del Chi-quadrato forniscono  $p$ -values ragionevolmente simili a quelli esatti: nell'esempio esaminato i  $p$ -values ottenuti attraverso il  $X^2$  ed il  $X_C^2$

risultano rispettivamente più piccoli e più grandi rispetto a quelli ottenuti con il test delle probabilità esatte di Fisher.

Tutti e tre i test parziali forniscono risultati significativi ( $p < 0.05$ ) per gli stessi due loci (*sAAT-4* e *bGALA-2*).

*Osservazione 7.3:* Questa affermazione di Ryman e Jorde non tiene in considerazione che la probabilità di commettere *almeno* un errore di I tipo (cioè di rifiutare almeno una sotto-ipotesi) quando è vera l'ipotesi nulla globale  $H_0$  e nel caso di indipendenza dei test parziali risulta:

$$\begin{aligned} \Pr\{\text{rifiutare almeno una delle } k \text{ sotto - ipotesi} \mid H_0 \cap \text{indip.}\} &= 1 - (1 - \alpha)^k \\ &= 1 - (1 - 0.05)^{12} = 0.4596 \end{aligned}$$

che rappresenta l'errore di I tipo multivariato, il *Familywise Error Rate* (FWE). Per l'inferenza marginale si necessita quindi una correzione per la molteplicità. Per una discussione maggiormente approfondita si veda il paragrafo 7.10.

### 7.3.1 SUMMATION METHOD

Si osservi che per la proprietà additiva del  $\chi^2$ , la somma di un insieme di variabili indipendenti  $\chi^2$ -distribuite segue una distribuzione  $\chi^2$  con un numero di gradi di libertà (g.d.l.) pari alla somma dei g.d.l. delle variabili contribuenti. Pertanto, per le statistiche Chi-quadrato ( $X^2$  e  $X_C^2$ ), che sotto l'ipotesi nulla sono asintoticamente  $\chi^2$ -distribuite con un grado di libertà, questa sommatoria equivale ad una combinazione diretta, cioè vengono sommati i valori delle  $k$  statistiche test parziali e il risultato viene valutato da una distribuzione  $\chi^2$  con  $k$  gradi di libertà.

Per i  $p$ -values ottenuti con i test esatti, Fisher (1950) dimostra che quando l'ipotesi nulla è vera, la quantità  $-2 \cdot \ln(\text{exact } p)$  segue asintoticamente una distribuzione  $\chi^2$  con due gradi di libertà. Quindi sommando il negativo di due volte il logaritmo naturale dei  $k$   $p$ -values ottenuti, il risultato viene valutato da una distribuzione  $\chi^2$  con  $2 \cdot k$  gradi di libertà (questa procedura viene denominata “metodo di Fisher”).

Per ciascuna delle tre statistiche test adottate ( $X^2$ ,  $X_C^2$  e il test delle probabilità esatte di Fisher) per valutare le  $k$  ( $= 12$ ) tabelle di contingenza, il *summation method* risulta significativo ( $p < 0.05$ ) e pertanto viene rifiutata l'ipotesi nulla globale (tabella 7.3).



### 7.3.2 METODO DI BONFERRONI

La logica di Bonferroni implica che nessun test parziale in un insieme di  $k$  test debba essere giudicato significativo a meno che il  $p$ -value considerato sia inferiore a  $\alpha/k$ , dove  $\alpha$  è il livello di significatività assegnato per rifiutare l'ipotesi nulla. Nel presente caso  $k = 12$  e  $\alpha = 0.05$ , pertanto viene richiesto un  $p$ -value inferiore a  $0.05 / 12 = 0.0042$  affinché esso venga considerato significativo.

Contrariamente al *summation-method*, l'applicazione del metodo di Bonferroni conduce ad una conclusione globale differente. Nel problema esaminato non c'è nessun  $p$ -value inferiore a 0.0042 e di conseguenza nessun locus può essere considerato tale da rilevare significativamente l'eterogeneità. L'ipotesi nulla globale  $H_0$  viene pertanto accettata per ciascuna delle tre statistiche test adottate.

In questo esempio il *summation method* porta al rifiuto dell'ipotesi nulla globale mentre il metodo di Bonferroni ne suggerisce l'accettazione, sebbene vengano considerati gli stessi insiemi di  $p$ -values, quindi indipendentemente dalla statistica test utilizzata.

La differenza è cruciale, ma in base ai dati disponibili un ricercatore non può determinare quale sia la decisione maggiormente appropriata. Sembra che molte pubblicazioni scientifiche abbiano accettato l'uno o l'altro metodo senza richiedere un'analisi dei dati supplementare.

## 7.4 CONCLUSIONI DI RYMAN E JORDE

E' evidente che la scelta del metodo statistico può essere fondamentale per trarre la giusta conclusione quando si esamina la differenziazione genetica.

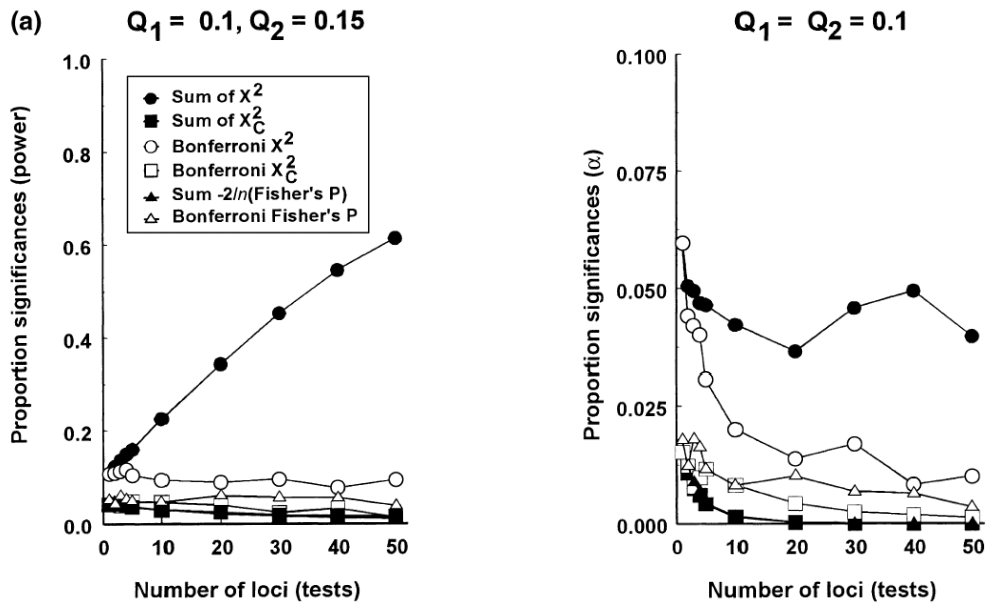
Sembra che l'interesse statistico principale sia stato dedicato a evitare le false significatività (errori di I tipo), mentre meno attenzione è stata prestata nella rilevazione delle vere differenze (potenza).

Poiché le due quantità sono correlate, l'attenzione eccessiva su una di esse può avere, d'altro canto, effetti poco desiderabili.

Sembra che la tendenza corrente in molti campi di biologia evolutiva sia di rilevare un numero costantemente crescente di loci in un numero abbastanza limitato di individui (nel *range*, per esempio, di 5-30 individui). Di conseguenza, il problema di come combinare le informazioni di più loci sta diventando sempre più importante.

Ryman e Jorde presentano i risultati di uno studio di simulazione valutando la probabilità di rilevazione di una effettiva differenza genetica (potenza statistica) e la probabilità di rifiutare erroneamente l'ipotesi nulla globale ( $\alpha$ -errore o errore di I tipo), confrontando il metodo di Bonferroni con *summation method*.

**Figura 6.1:** Potenza statistica e errore di I tipo per il *summation method* e il metodo di Bonferroni applicati ai test  $X^2$ ,  $X_C^2$  e al test esatto di Fisher; la frequenza allelica per il primo e il secondo campione è pari rispettivamente a  $Q_1 = 0.10$  e  $Q_2 = 0.15$ ;  $\alpha = 0.05$ ;  $n_1 = n_2 = 20$ ; il numero di simulazioni Monte Carlo è pari a 10 000 (Ryman e Jorde, 2001).



#### 7.4.1 SUMMATION METHOD: SOMMA DI STATISTICHE CHI-QUADRATO

I risultati indicano che il *summation method* applicato ai test parziali ottenuti attraverso il Chi-quadrato convenzionale ( $X^2$ ) tende ad dare risultati migliori rispetto alle altre due alternative esaminate (il  $X_C^2$  e il test delle probabilità esatte di Fisher), anche per numerosità campionarie ragionevolmente piccole (figura 6.1).

La logica sottostante al *summation method* si basa sul fatto che la somma di due o più variabili  $\chi^2$ -distribuite e indipendenti segue una distribuzione  $\chi^2$  con gradi di libertà pari alla somma dei gradi di libertà delle variabili componenti. L'adattamento alla distribuzione  $\chi^2$  può comunque risultare abbastanza scarso, particolarmente per numerosità campionarie basse, e la somma di parecchie variabili "scarsamente adattate" può allontanarsi drasticamente dalla distribuzione  $\chi^2$  attesa.

La media e la varianza della distribuzione  $\chi^2$  sono pari a g.d.l. e  $2 \cdot \text{g.d.l.}$  rispettivamente, e come valutazione dell'adattamento alla distribuzione teorica  $\chi^2$  la tabella 7.4 riporta la media e la varianza osservate delle statistiche test ottenute simulando di estrarre due campioni indipendenti da una popolazione con frequenza effettiva dell'allele raro pari a 0.1 e uguale numerosità campionaria  $n$ . Il numero di ripetizioni (numero di tabelle  $2 \times 2$ ) è pari a 10 000 per ogni numerosità campionaria considerata.

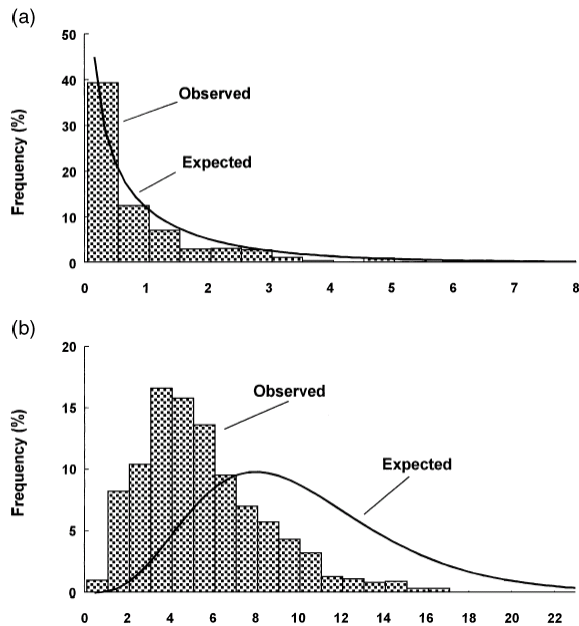
<i>statistica test</i>	<i>n</i>	<i>media</i>		<i>varianza</i>	
		<i>osservata</i>	<i>attesa</i>	<i>osservata</i>	<i>attesa</i>
$X^2$	3	1.09	1	0.75	2
$X^2$	10	1.04	1	1.52	2
$X^2$	20	1.01	1	1.77	2
$X^2$	50	1.00	1	1.95	2
$X^2$	100	1.02	1	2.04	2
$X^2$	500	1.00	1	1.98	2
$X_C^2$	3	0.25	1	0.15	2
$X_C^2$	10	0.44	1	0.49	2
$X_C^2$	20	0.54	1	0.83	2
$X_C^2$	50	0.67	1	1.28	2
$X_C^2$	100	0.78	1	1.53	2
$X_C^2$	500	0.88	1	1.75	2
$-2 \cdot \ln(\text{exact } p)$	3	0.34	2	0.59	4
$-2 \cdot \ln(\text{exact } p)$	10	0.84	2	1.74	4
$-2 \cdot \ln(\text{exact } p)$	20	1.13	2	2.47	4
$-2 \cdot \ln(\text{exact } p)$	50	1.41	2	3.15	4
$-2 \cdot \ln(\text{exact } p)$	100	1.59	2	3.49	4
$-2 \cdot \ln(\text{exact } p)$	500	1.80	2	3.70	4

**Tabella 7.4:**

Media e varianza delle statistiche test per tabelle di contingenza 2x2 simulando di estrarre due campioni indipendenti di uguale numerosità ( $n = 3 - 500$ ) da una popolazione con frequenza effettiva dell'allele raro pari a 0.1. Il numero di ripetizioni (numero di tabelle 2x2) è pari a 10 000 per ogni numerosità campionaria considerata (Ryman e Jorde, 2001).

L'adattamento della statistica test  $X^2$  alla distribuzione teorica  $\chi^2$  è abbastanza buono, ad eccezione per numerosità campionarie basse, per le quali la varianza risulta ridotta e la media aumentata.

Al contrario, l'adattamento di  $X_C^2$  e di  $-2 \cdot \ln(\text{exact } p)$  alla distribuzione limite  $\chi^2$  è più lento e la distribuzione campionaria osservata tende a localizzarsi alla sinistra di quella attesa; questo spostamento di posizione produce un numero di false significatività inferiore a quello progettato ( $\alpha$ -realizzato < 0.05) e pertanto la potenza viene ridotta. Questo spostamento di posizione risulta maggiormente evidente quando si considera la somma (diretta o del negativo di due volte il logaritmo del *p-value*) delle singole statistiche test nel confronto con la distribuzione teorica attesa per la loro somma (figura 6.2):



**Figura 6.2:** Simulazione della distribuzione campionaria della statistica test  $\chi^2_C$  quando si estraggono due campioni indipendenti di 20 unità (40 geni) da una popolazione con frequenza allelica pari a 0.10. Le corrispondenti distribuzioni  $\chi^2$  sono quelle attese asintoticamente nella teoria di campioni numerosi.

- (a) si esamina un locus singolo; il  $\chi^2$  atteso ha un grado di libertà;
- (b) si verificano 10 loci separatamente e si sommano il valori delle statistiche test; il  $\chi^2$  atteso ha dieci gradi di libertà;

Dai risultati ottenuti, Ryman e Jorde concludono che il *summation-method* applicato alle  $k$  statistiche test Chi-Quadrato convenzionali dovrebbe rappresentare tipicamente il metodo da utilizzare per verificare l'ipotesi nulla globale  $H_0$ .

#### 7.4.2 SUMMATION METHOD: METODO DI FISHER

Contro la tendenza di impiegare sempre più spesso il “metodo di Fisher” applicato ai  $p$ -values esatti Ryman e Jorde mettono in evidenza che a questo metodo può essere associata una potenza bassa anche quando le dimensioni del campione sono abbastanza grandi. Per effettuare il test complessivo può sembrare superficiale trovare allettante che le probabilità sono state calcolate “esattamente” per ogni singola tabella di contingenza componente. Tuttavia, lo scarso adattamento alla distribuzione  $\chi^2$  prevista asintoticamente spesso rende questo metodo abbondantemente conservativo.

Si vuole evidenziare che Ryman e Jorde non suggeriscono che i test esatti di contingenza debbano essere abbandonati. Non esiste un test per tabelle di contingenza che sia universalmente il migliore per ogni circostanza e vi sono molte situazioni nelle quali un ricercatore può avere la valida preoccupazione per quanto riguarda la convenienza dell'utilizzo di una statistica Chi-quadrato, che necessariamente rappresenta un'approssimazione. Con valori attesi bassi in una o più celle il rischio di ottenere false significatività non può essere ignorato, anche se parecchi studi suggeriscono che la gravità di questo problema di continuità potrebbe essere stato sopravvalutato.

Per quanto possibile, essi affermano, il calcolo esatto del  $p$ -value dovrebbe essere preferito all'approssimazione che conta sulla teoria di un campione numeroso, in particolare quando le dimensioni del campione sono modeste o piccole. Anche se risultano conservativi, è un ovvio vantaggio condurre test esatti in quanto il ricercatore ha la

garanzia che l' $\alpha$  realizzato non eccederà mai quello progettato. Il problema riguarda quindi la *combinazione* delle informazioni di più test esatti per mezzo di un'approssimazione quale il metodo di Fisher.

#### 7.4.3 METODO DI BONFERRONI

La correzione di Bonferroni è stata progettata principalmente per ridurre la probabilità di ottenere false significatività quando si effettuano test multipli, non è mirata invece a combinare le informazioni di più test parziali per verificare la stessa ipotesi nulla<sup>2</sup>. Il metodo di Bonferroni si concentra esclusivamente sul caso di  $p$ -values molto piccoli e in gran parte ignora tendenze deboli di significatività, mentre qualsiasi indicazione sulle differenze di frequenza dell'allele in qualsiasi locus polimorfo (indipendentemente dalla direzione) dovrebbe contribuire idealmente con delle informazioni che rendano l'ipotesi nulla globale meno probabile<sup>3</sup>.

Affinché il metodo di Bonferroni sia di interesse pratico, devono essere esaudite due condizioni di base che in molte situazioni reali non vengono soddisfatte:

- 1) Per prima cosa è necessario che ogni test parziale possa produrre un  $p$ -value che sia inferiore ad  $\alpha/k$ ; spesso però ciò non è possibile per molte combinazioni di numerosità campionarie e di frequenze dell'allele nella popolazione. Si consideri ad esempio di campionare due individui (e quindi quattro geni) da ognuna delle due popolazioni. Si supponga che i conteggi dei due tipi di alleli risultino pari a 4; 0 e 0; 4 per i due campioni rispettivamente, e ciò rappresenta il risultato più estremo possibile sotto l'ipotesi nulla di frequenze uguali. Il test delle probabilità esatte di Fisher produce un  $p$ -value (bidirezionale) pari a 0.029, significativo ad un livello  $\alpha$  pari a 0.05, e rappresenta il  $p$ -value più piccolo che può essere ottenuto con queste numerosità campionarie. Se analizziamo cinque loci che rendono lo stesso risultato di quello sopra descritto, intuitivamente sarebbe legittimo attendersi che le popolazioni siano geneticamente differenti. Una valutazione attraverso il metodo di Bonferroni scarterebbe una tale interpretazione perché nessun  $p$ -value è inferiore a  $0.05 / 5 = 0.01$ , infatti risulta impossibile ottenere un  $p$ -value così piccolo con le dimensioni campionarie considerate. Così, in questa situazione, la correzione di Bonferroni produce un tasso di errore di I tipo pari a zero e pertanto anche potenza viene ridotta a zero.
- 2) Si richiede inoltre che l'errore  $\alpha$  realizzato per ogni singolo locus sia ragionevolmente vicino a quello inteso. Quando l' $\alpha$  realizzato nel singolo test di contingenza è inferiore a quello progettato, allora l'approccio di Bonferroni finalizzato a verificare l'ipotesi nulla globale può risultare conservativo. Per

<sup>2</sup> Si veda capitolo 6, paragrafo 6.3, punto (3);

<sup>3</sup> Si può parlare rispettivamente di deviazioni "pure" o "miste": si veda il capitolo 6, paragrafi 6.2 e 6.7;

esemplificare ciò, la tabella 7.5 mostra i valori esatti dell' $\alpha$  realizzato per 1-50 loci quando si applica il test delle probabilità esatte di Fisher su due campioni indipendenti di numerosità  $n = 10, 20$  da una popolazione con frequenza effettiva dell'allele raro pari a 0.10:

$k =$ numero di loci	$p$ -value inteso per il rifiuto	$n = 10$		
		$p$ -value realizzato per rifiutare una singola tabella $2 \times 2$	frequenza di accadimento	$\alpha$ realizzato
1	0.05	0.04837	0.012002	0.012002
2	0.025	0.02484	0.003010	0.006010
5	0.01	0.00953	0.000622	0.003107
10	0.005	0.00385	0.000107	0.001065
20	0.0025	0.00220	0.000015	0.000304
50	0.001	0.00077	0.000002	0.000090

$k =$ numero di loci	$p$ -value inteso per il rifiuto	$n = 20$		
		$p$ -value realizzato per rifiutare una singola tabella $2 \times 2$	frequenza di accadimento	$\alpha$ realizzato
1	0.05	0.04817	0.016770	0.016770
2	0.025	0.02470	0.005559	0.011087
5	0.01	0.00931	0.002022	0.010067
10	0.005	0.00483	0.000678	0.006761
20	0.0025	0.00240	0.000514	0.010225
50	0.001	0.00096	0.000056	0.002802

**Tabella 7.5:**  
 $\alpha$ -errori realizzati quando si applica il metodo di Bonferroni a una serie di  $k$  tabelle di contingenza  $2 \times 2$ . Ogni tabella rappresenta due campioni indipendenti di  $n$  individui diploidi ( $2 \cdot n$  geni) da una popolazione dove la frequenza effettiva dell'allele raro è pari a 0.10, e ad essa viene applicato il test delle probabilità esatte di Fisher. La frequenza di accadimento rappresenta la probabilità binomiale cumulata di ottenere il  $p$ -value realizzato per il rifiuto o uno inferiore ad esso. (Ryman e Jorde, 2001)

Per esempio, per  $n = 10$  e  $k = 10$ , il metodo di Bonferroni implica che l'ipotesi nulla globale viene rifiutata se almeno una tabella di contingenza produce un  $p$ -value inferiore a 0.005: il  $p$ -value di Fisher più grande che soddisfi a questo criterio è pari a 0.00385 e un  $p$ -value così piccolo o inferiore accade con una frequenza di 0.000107 (probabilità binomiale cumulata). Quindi, l' $\alpha$  realizzato dall'approccio di Bonferroni per verificare  $H_0$  corrisponde alla probabilità di osservare uno o più  $p$ -values  $\leq 0.05$ , e cioè  $\alpha = 1 - (1 - 0.000107)^{10} = 0.001065$ , che è notevolmente inferiore a quello progettato.

## 7.5 SOLUZIONE DI PERMUTAZIONE

L'intero set di dati  $\mathbf{X}$  è un insieme di statistiche sufficienti per il problema in esame, perciò occorre costruire  $k$  test parziali ( $h = 1, \dots, k$ ) che soddisfino ai rispettivi e distinti sistemi d'ipotesi:

$$H_{0h} : \{X_{h1} \stackrel{d}{=} X_{h2}\} \text{ contro } H_{1h} : \{X_{h1} \stackrel{d}{\neq} X_{h2}\} \quad \text{per } h = 1, \dots, k$$

7.5.1 TEST DI PERMUTAZIONE PARZIALI

Attraverso la teoria standard dei test di permutazione, si ottengono i  $k$  test parziali

$$T_h^* = (\sum_{i=1}^{2 \cdot n_2} X_{h2i}^* / 2 \cdot n_2) - (\sum_{i=1}^{2 \cdot n_1} X_{h1i}^* / 2 \cdot n_1)$$

dove  $h = 1, \dots, k$ , ed i soggetti si intendono permutati (secondo una procedura casuale) all'interno dell'insieme totale di  $2 \cdot n = 2 \cdot n_1 + 2 \cdot n_2$  osservazioni:

$$\mathbf{X}_h = \mathbf{X}_{h1} \oplus \mathbf{X}_{h2} \Rightarrow \mathbf{X}_h^* = (\mathbf{X}_{h1} \oplus \mathbf{X}_{h2})^*$$

*Osservazione 7.4:* Le risposte  $X_{hji}$  sono binarie:  $X_{hji} = 0$  o  $1, i = 1, \dots, 2 \cdot n_j, j = 1, 2, h$

$= 1, \dots, k$ . Se per esempio, l'alternativa  $H_1$  risulta  $H_1 : \{\cup_{h=1}^k [X_{h1} <^d X_{h2}]\}$  un test parziale unidirezionale di permutazione appropriato risulta essere  $T_h^* = \sum_{i=1}^{2 \cdot n_2} X_{h2i}^*$

(permutazionalmente equivalente a  $T^*$ ). Si noti che per  $k = 1$  questo problema corrisponde al test delle probabilità esatte di Fisher, nel senso che le sotto-ipotesi vengono verificate da test delle probabilità esatte di Fisher parziali. Se  $k > 1$  otteniamo una naturale estensione multivariata del test esatto di Fisher, che può effettivamente essere risolta con il metodo di combinazione non parametrica. Comunque, vale la pena notare che la distribuzione multivariata di  $(T_1^*, \dots, T_k^*)$  è tale che ogni componente marginale univariata è ipergeometrica, mentre la distribuzione multivariata non è multivariata ipergeometrica.

Quindi per stimare la distribuzione di permutazione della statistica test  $T_h^*, h = 1, \dots, k$ , si utilizzano  $B$  ricampionamenti casuali senza reinserimento, per cui è possibile riassumere la procedura di ricampionamento nella seguente tabella:

**Tabella 7.6:** Procedura CMC per il test  $T_h$

$\mathbf{X}_h = (\mathbf{X}_{h1} \oplus \mathbf{X}_{h2})$	$(\mathbf{X}_{h1} \oplus \mathbf{X}_{h2})_1^*$	...	$(\mathbf{X}_{h1} \oplus \mathbf{X}_{h2})_B^*$
$T_h^{oss}$	${}_1T_h^*$	...	${}_B T_h^*$

Nelle verifiche multivariate, in modo da preservare tutte le relazioni di dipendenza sottostanti fra le variabili, le permutazioni *devono essere sempre* eseguite sui singoli vettori di dati, in modo che tutte le variabili componenti ed i test parziali *siano analizzati congiuntamente*. Il concetto di test marginali può talvolta indurre erroneamente a pensare che si possa considerare più permutazioni indipendenti per ogni test componente, ma in questo modo i rapporti di dipendenza multivariati possono non essere preservati.

Dato quindi un livello di significatività  $\alpha$ , si accetterà l'ipotesi nulla parziale  $H_{0h}$  se il  $p$ -value di permutazione stimato sulla statistica  $T_h^*$ , ed indicato con:

$$\hat{\lambda}_h = \#(T_h^* \geq T_h^{oss}) / B \notin \begin{cases} [0, 1 - \alpha] & \Rightarrow H_{1h} : \{X_{h1} <^d X_{h2}\} \\ (\alpha/2, 1 - \alpha/2) & \Rightarrow H_{1h} : \{X_{h1} \neq^d X_{h2}\} \\ (\alpha, 1] & \Rightarrow H_{1h} : \{X_{h1} >^d X_{h2}\} \end{cases}$$

dove  $B$  è il numero di permutazioni considerate nella simulazione.

Le statistiche test parziali si dimostrano essere permutazionalmente esatte (le variazioni delle statistiche sotto l'ipotesi nulla dipendono solo dalla casualità delle permutazioni), non distorte e consistenti. Essendo i test parziali marginalmente non distorti e significativi per valori elevati oppure per valori bassi, allora le loro distribuzioni di permutazione sotto  $H_1$  risultano stocasticamente superiori o inferiori a quelle sotto  $H_0$ . Ciò significa che

$$\forall \alpha > 0, h = 1, \dots, k \begin{cases} \Pr\{\hat{\lambda}_h \geq (1 - \alpha) \mid H_{1h} : [X_{h1} <^d X_{h2}]\} \geq \alpha \\ \Pr\{\alpha/2 < \hat{\lambda}_h < (1 - \alpha/2) \mid H_{1h} : [X_{h1} \neq^d X_{h2}]\} \leq 1 - \alpha \\ \Pr\{\hat{\lambda}_h \leq \alpha \mid H_{1h} : [X_{h1} >^d X_{h2}]\} \geq \alpha \end{cases}$$

La consistenza dei test parziali invece implica che

$$\forall \alpha > 0, h = 1, \dots, k \begin{cases} \lim_{n \rightarrow \infty} \Pr\{\hat{\lambda}_h < (1 - \alpha) \mid H_{1h} : [X_{h1} <^d X_{h2}]\} = 0 \\ \lim_{n \rightarrow \infty} \Pr\{\alpha/2 < \hat{\lambda}_h < (1 - \alpha/2) \mid H_{1h} : [X_{h1} \neq^d X_{h2}]\} = 0 \\ \lim_{n \rightarrow \infty} \Pr\{\hat{\lambda}_h > \alpha \mid H_{1h} : [X_{h1} >^d X_{h2}]\} = 0 \end{cases}$$

Inoltre i test  $\lambda_h, h = 1, \dots, k$  sono monotonicamente associati sotto  $H_1$  (Pesarin, 2001).

A questo punto, la soluzione di permutazione utilizza un metodo di combinazione non parametrica sui test parziali. Questa combinazione in un unico test di secondo ordine  $\mathbf{T}'' = \psi(\lambda_1, \dots, \lambda_k)$  è ottenuta con una funzione reale (misurabile) continua, monotona decrescente e non degenera

$\psi : (0, 1)^k \rightarrow \mathfrak{R}^1$ . Per essere una buona funzione di combinazione,  $\psi$  deve soddisfare le proprietà (P.1), (P.2) e (P.3) presentate nel capitolo 6, paragrafo 6.5.2; per l'algoritmo per la simulazione della procedura di combinazione non parametrica si veda il paragrafo 6.5.3.



7.5.2 PROCEDURA MULTI-ASPECT

Ryman e Jorde (2001) sebbene abbiano premesso che “nel mondo reale non è molto verosimile incontrare una situazione nella quale tutti i loci esaminati in una popolazione forniscano esattamente la stessa frequenza”, hanno simulato solamente situazioni in cui tutti i loci all'interno di una popolazione hanno frequenze dell'allele identiche. Per tali situazioni risulta che il *summation method* è più potente della correzione di Bonferroni (figura 6.1).

E' ben noto che la mutabilità del DNA varia sostanzialmente tra i siti in un genoma (Foster *et al.*, 1982). È *a priori* non noto se vi sono differenze per quanto riguarda uno, alcuni o quasi tutti i loci (o nessuno nel caso in cui l'ipotesi nulla globale risulti vera).

A questo fine, abbiamo bisogno di un test statistico robusto, cioè di un test che sia relativamente potente per tutte le possibili situazioni sotto l'ipotesi alternativa, valere a dire sia quando esistono soltanto una o poche differenze sia quando esistono varie differenze (anche se “deboli”, ma che possano venir accumulate) tra i  $k$  loci esaminati.

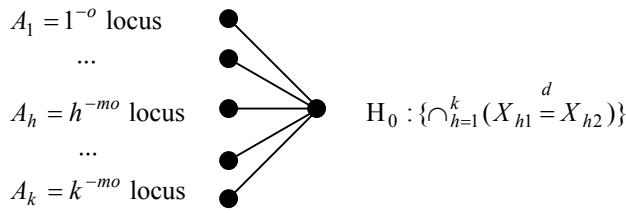
Seguendo la procedura generale multi-aspect descritta nel capitolo 6, paragrafo 6.4, si procede in tre fasi. La prima fase, in accordo con l'osservazione 7.2, è subito soddisfatta: gli aspetti per il problema sono rappresentati dai  $k$  loci considerati (tabella 7.7a).

**Tabella 7.7a:** Prima fase

<i>ipotesi nulla</i>	<i>aspetti considerati</i>	<i>statistiche test</i>	<i>sotto-ipotesi</i>
$H_0 : \{\bigcap_{h=1}^k (X_{h1}^d = X_{h2}^d)\}$ contro $H_1 : \{\bigcup_{h=1}^k (X_{h1}^d \neq / < X_{h2}^d)\}$	$A_1 = 1^{-o}$ locus	$T_1 / \overset{P}{T}_1$	$H_{11} : \{X_{11}^d \neq X_{12}^d\} / H_{11} : \{X_{11}^d < X_{12}^d\}$
	...	...	...
	$A_h = h^{-mo}$ locus	$T_h / \overset{P}{T}_h$	$H_{1h} : \{X_{h1}^d \neq X_{h2}^d\} / H_{1h} : \{X_{h1}^d < X_{h2}^d\}$
	...	...	...
	$A_k = k^{-mo}$ locus	$T_k / \overset{P}{T}_k$	$H_{1k} : \{X_{k1}^d \neq X_{k2}^d\} / H_{1k} : \{X_{k1}^d < X_{k2}^d\}$

dove il simbolo ‘ / ’ separa rispettivamente il caso bidirezionale da quello unidirezionale, mentre i test parziali di permutazione corrispondono a quelli descritti nel paragrafo 7.5.1.

Nella seconda fase (figura 6.3), si struttureranno gli aspetti “direttamente”: i locus considerati e i relativi test parziali sono infatti “omogenei”, e in mancanza di ulteriori informazioni non si può procedere altrimenti. Così facendo, si assegnerà la medesima importanza ai  $k$  loci considerati.



**Figura 7.3:** Seconda fase della procedura multi-aspect

Nella terza fase (ignorandone il primo passo), in virtù delle considerazioni di inizio paragrafo, si opererà per una forma di combinazione neutra, la procedura di combinazione iterata, nel seguente modo (tabella 7.7b-ii):

**Tabella 7.7b-ii:** Terza fase, secondo passo: procedura di combinazione iterata

n° di loci (su $k$ ) sotto $H_1$ (●)	funzione di combinazione “appropriata”	funzione di combinazione “robusta”
●○...○○...○○	⇒ funzione di combinazione Tippett	} funzione di combinazione iterata
●●...●○...○○	⇒ funzione di combinazione Fisher	
●●...●●...●●	⇒ funzione di combinazione Liptak <sup>4</sup>	

Infatti, in base alle considerazioni del paragrafo 6.7, si dovrebbe scegliere rispettivamente la funzione di combinazione di Tippett, di Fisher o di Liptak se pochi, abbastanza o molti loci sono attesi presentare eterogeneità; in mancanza di tali informazioni a *priori*, in accordo alle osservazioni del paragrafo 6.4, la procedura iterata risulta *robusta* rispetto alla scelta delle funzioni di combinazione.

### 7.6 STUDIO COMPARATIVO VIA SIMULAZIONE

Si considererà il caso in cui tutti i loci presso una popolazione presentano frequenze identiche dell’allele raro. Sebbene questa situazione non sia molto verosimile, tuttavia la presentiamo sia per esigenze comparative con lo studio di Ryman e Jorde (2001), sia perché gli stessi autori argomentano questa scelta affermando che, volendo esaminare le differenze fra i metodi di test, dal momento che sia la potenza sia l’ $\alpha$  errore dipendono dalle frequenze della popolazione, il variare di queste frequenze potrebbe oscurarne il confronto. Pertanto

<sup>4</sup> si noti che, in condizioni di regolarità (si veda Pesarin, 2001) che conducono a una forma di verifica parametrica asintoticamente ottimale, se tutte le  $k$  alternative sono congiuntamente vere con lo stesso parametro di non-centralità standardizzato (così che, quando considerati marginalmente, tutti i test parziali sono ugualmente potenti), il test combinato di Liptak  $T_L''$  è asintoticamente equivalente al migliore test parametrico;

$$\forall k \in \mathbb{N}_+, \Pr\{Y_{hj} = A_h\} = \Pr\{X_{hj} = 1\} = \Pr\{X_j = 1\} = Q_j; \quad h = 1, \dots, k, \quad j = 1, 2.$$

Per ciascuna popolazione  $j$ -esima ( $j = 1, 2$ ) verranno quindi generate  $2 \cdot n_j$  ( $n_j$  individui diploidi) prove bernoulliane con probabilità di successo pari a  $Q_j$  per ogni locus  $h$ -esimo considerato ( $h = 1, \dots, k$ ). L'assunto che gli alleli all'interno degli individui sono indipendenti è giustificato dall'equilibrio di Hardy-Weinberg.

Le seguenti simulazioni Monte Carlo saranno effettuate con un livello di significatività  $\alpha = 0.05$ , un numero di simulazioni Monte Carlo  $MC = 2000$  e un numero di permutazioni (ricampionamenti condizionati CMC)  $B = 1000$ . La numerosità campionaria sarà uguale per i due campioni ( $n_1 = n_2$ ) e sarà pari a 20 (40 geni), con  $Q_1 = 0.1$  e  $Q_2 = Q_1$  sotto  $H_0$  e  $Q_2 = 0.15$  sotto  $H_1$ .

Per ogni procedura di verifica, nella figura 7.2 vengono indicate le proporzioni di ripetizioni ottenute che rifiutano l'ipotesi nulla globale  $H_0$  (asse  $y$ ) per differenti quantità di loci esaminati (asse  $x$ ,  $k = 1, 2, \dots, 5, 10, 20, \dots, 50$ ). Quando  $Q_1$  è differente da  $Q_2$  ( $H_0$  falsa) la proporzione di rifiuti di  $H_0$  stima la potenza del test, mentre quando  $Q_1 = Q_2$  ( $H_0$  vera) stima l'errore di I tipo realizzato.

In un primo passo, i test parziali usati per verificare l'omogeneità di frequenza dell'allele raro in ogni locus separatamente sono i seguenti (tabella 7.8):

**Tabella 7.8:** Test parziali per verificare l'omogeneità di frequenza nel locus  $h$ -esimo

<i>test parziale</i>	<i>espressione del test</i>	<i>sotto-ipotesi alternativa</i>
$X_h^2$	$\sum_{j=1,2} \sum_{h=A_h, a_h} (f_{hj} - E[f_{hj}])^2 / E[f_{hj}] \sim \chi_1^2$	$H_{1h} : \{X_{h1} \stackrel{d}{\neq} X_{h2}\}$
$T_h^*$	$(\sum_{i=1}^{2 \cdot n_2} X_{h2i}^* / 2 \cdot n_2) - (\sum_{i=1}^{2 \cdot n_1} X_{h1i}^* / 2 \cdot n_1)$	$H_{1h} : \{X_{h1} \stackrel{d}{\neq} X_{h2}\}$
$T_h^{p*}$	$\sum_{i=1}^{2 \cdot n_2} X_{h2i}^* / 2 \cdot n_2$	$H_{1h} : \{X_{h1} \stackrel{d}{<} X_{h2}\}$

dove  $f_{hj}$  è la frequenza osservata e  $E[f_{hj}] = f_h \cdot n_j / (n_1 + n_2)$  quella attesa;  $j = 1, 2$ ,  $h = A_h, a_h$ . Viene considerato il test unidirezionale  $T_h^{p*}$  poiché quando è vera  $H_1$  abbiamo  $Q_2 > Q_1$ .

In un secondo passo, vengono combinati i  $k$  test parziali in un test globale nel seguente modo (tabella 7.9):

**Tabella 7.9:** Combinazione dei  $k$  test parziali nel test globale

<i>test globale</i>	<i>funzioni di combinazione</i>	<i>ipotesi alternativa globale</i>
$\sum_{h=1}^k X_h^2$	$\sum_{h=1}^k X_h^2 \sim \chi_k^2$	$H_1 : \{\cup_{h=1}^k [X_{h1} \stackrel{d}{\neq} X_{h2}]\}$
$T_l''$	$\psi_l(\lambda_1, \dots, \lambda_k) : l = \text{Tippett, Fisher, Liptak}$	$H_1 : \{\cup_{h=1}^k [X_{h1} \stackrel{d}{\neq} X_{h2}]\}$
$T_l''^U$	$\psi_l(\lambda_1^U, \dots, \lambda_k^U) : l = \text{Tippett, Fisher, Liptak}$	$H_1 : \{\cup_{h=1}^k [X_{h1} \stackrel{d}{<} X_{h2}]\}$
$T_l'''$	$\psi_l(\lambda_T'', \lambda_F'', \lambda_L'') : l = \text{Fisher}$	$H_1 : \{\cup_{h=1}^k [X_{h1} \stackrel{d}{\neq} X_{h2}]\}$
$T_l'''^U$	$\psi_l(\lambda_T''^U, \lambda_F''^U, \lambda_L''^U) : l = \text{Fisher}$	$H_1 : \{\cup_{h=1}^k [X_{h1} \stackrel{d}{<} X_{h2}]\}$

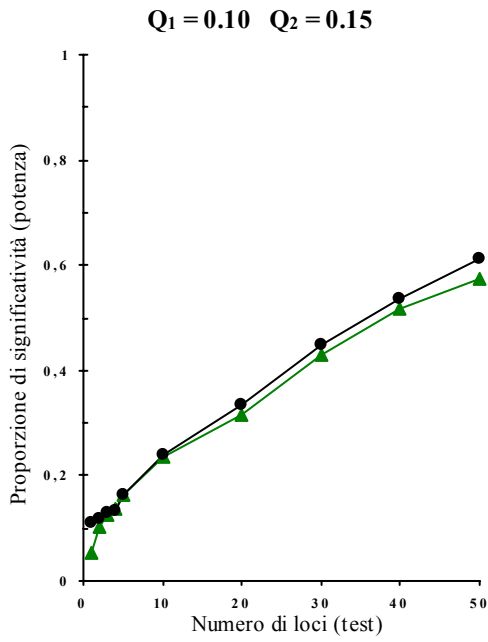
Si è scelto di presentare i risultati del test globale  $T_l'''$  ( $T_l'''^U$ ) poiché esso verrà utilizzato nel caso più generale in cui le differenze riguardano  $z \leq k$  loci (paragrafo 7.7); si presenteranno inoltre i risultati parziali per le funzioni di combinazione  $T_l''$  ( $T_l''^U$ ) con  $l = \text{Tippett, Fisher, Liptak}$ . I risultati vengono presentati nella tabella 7.10 e nelle figure 7.4a e 7.4b: si può osservare, per quanto riguarda la stima della potenza, che la soluzione iterata risulta approssimativamente equivalente al *summation method* nel caso bidirezionale, mentre dà risultati migliori nel caso unidirezionale.

*Osservazione 7.5:* Quando tutte le statistiche test parziali sono omogenee, per cui condividono esattamente la stessa distribuzione di permutazione asintotica (per esempio, se sono tutte del tipo Chi-quadrato con lo stesso numero di gradi di libertà) e il loro supporto asintotico comune è almeno non limitato a destra, allora si può prendere in considerazione la funzione di combinazione non parametrica diretta, espressa nella forma  $T_D'' = \sum_{i=1}^k T_i$ ,  $T_{Do}'' = \sum_{i=1}^k T_{io}$ , e  $T_{Dr}''^* = \sum_{i=1}^k T_{ir}^*$ ,  $r = 1, \dots, B$  per il test combinato, osservato e per i valori di permutazione, rispettivamente. Il  $p$ -value combinato è dato da  $\hat{\lambda}'' = \sum_{r=1}^B \mathbf{I}(T_{Dr}''^* \geq T_{Do}'') / B$ . Si osservi che le distribuzioni di permutazione dei tutti i test parziali possono essere le stesse solo per numerosità campionarie abbastanza elevate. Quindi, per numerosità campionarie finite, questa condizione può essere soddisfatta approssimativamente poiché le distribuzioni di permutazione sono essenzialmente dipendenti dai dati osservati. La funzione di combinazione diretta ci permette di evitare i calcoli abbastanza intensivi coinvolti nei passi (S.b<sub>k</sub>) e (C.a) – (C.c) del capitolo 6, paragrafo 6.5.3. Inoltre, anche se la funzione di combinazione diretta opera come nel caso univariato, è essenzialmente una combinazione nonparametrica giacché: (i) il problema di verifica è equivalentemente spezzato in  $k$  sotto-problemi; (ii) una statistica test globale  $T : \mathfrak{R}^k \rightarrow \mathfrak{R}^1$  non è direttamente disponibile; (iii) le relazioni di dipendenza tra i test parziali sono implicitamente “catturate” dalla procedura di combinazione (Pesarin, 2001).

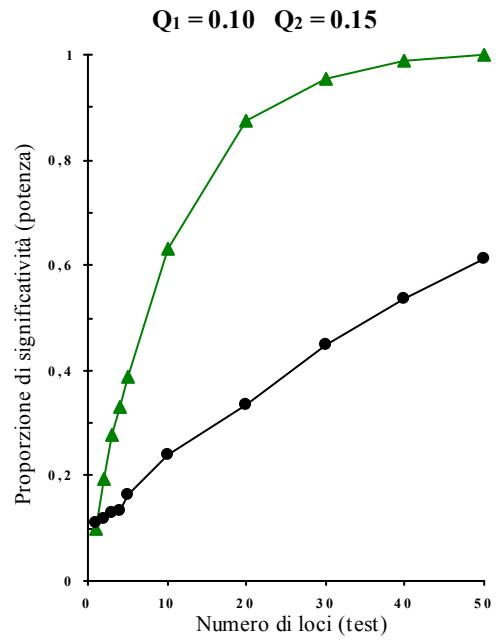
**Tabella 7.10:** Studio comparativo via simulazione

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)						
Q <sub>1</sub> = 0.10, Q <sub>2</sub> = 0.15; n <sub>1</sub> = n <sub>2</sub> = 20; α = 0.05; MC = 2000*, B = 1000*; k = 1,2,...,5,10,20,...,50;						
loci	$\psi_T(\lambda_1, \dots, \lambda_k)$	$\psi_F(\lambda_1, \dots, \lambda_k)$	$\psi_L(\lambda_1, \dots, \lambda_k)$	$\psi_T^p(\lambda_1, \dots, \lambda_k)$	$\psi_F^p(\lambda_1, \dots, \lambda_k)$	$\psi_L^p(\lambda_1, \dots, \lambda_k)$
1	.0545	.0545	.0545	.0975	.0975	.0975
2	.0720	.1030	.0910	.1235	.1940	.2035
3	.0850	.1275	.1165	.1470	.2695	.3010
4	.0990	.1390	.1160	.1580	.3325	.3630
5	.1105	.1645	.1315	.1905	.3900	.4325
10	.1380	.2440	.1885	.2325	.6285	.7055
20	.1413	.3393	.2520	.2313	.8613	.9147
30	.1340	.4470	.3590	.2160	.9530	.9740
40	.1500	.5430	.4210	.2350	.9840	.9930
50	.1140	.6110	.4720	.2080	.9980	1.000
loci	$\psi_F(\lambda_T^*, \lambda_F^*, \lambda_L^*)$	$\sum_{h=1}^k X_h^2 \sim \chi_k^2$	$\psi_F^p(\lambda_T^*, \lambda_F^*, \lambda_L^*)$			
1	.0545	.1090	.0975			
2	.1025	.1170	.1945			
3	.1255	.1305	.2760			
4	.1355	.1325	.3305			
5	.1620	.1640	.3870			
10	.2350	.2385	.6315			
20	.3167	.3360	.8747			
30	.4280	.4500	.9540			
40	.5170	.5350	.9870			
50	.5740	.6120	.9990			
H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)						
Q <sub>1</sub> = Q <sub>2</sub> = 0.10; n <sub>1</sub> = n <sub>2</sub> = 20 (40 geni); α = 0.05; MC = 2000*, B = 1000*; k = 1,2,...,5,10,20,...,50;						
loci	$\psi_T(\lambda_1, \dots, \lambda_k)$	$\psi_F(\lambda_1, \dots, \lambda_k)$	$\psi_L(\lambda_1, \dots, \lambda_k)$	$\psi_T^p(\lambda_1, \dots, \lambda_k)$	$\psi_F^p(\lambda_1, \dots, \lambda_k)$	$\psi_L^p(\lambda_1, \dots, \lambda_k)$
1	.0220	.0220	.0220	.0215	.0215	.0215
2	.0235	.0375	.0345	.0250	.0385	.0370
3	.0295	.0460	.0450	.0335	.0415	.0490
4	.0255	.0435	.0390	.0315	.0460	.0475
5	.0320	.0445	.0450	.0420	.0530	.0530
10	.0413	.0440	.0540	.0480	.0485	.0525
20	.0370	.0370	.0330	.0350	.0440	.0500
30	.0340	.0500	.0550	.0450	.0570	.0610
40	.0380	.0360	.0480	.0310	.0360	.0370
50	.0330	.0470	.0540	.0410	.0500	.0470
loci	$\psi_F(\lambda_T^*, \lambda_F^*, \lambda_L^*)$	$\sum_{h=1}^k X_h^2 \sim \chi_k^2$	$\psi_F^p(\lambda_T^*, \lambda_F^*, \lambda_L^*)$			
1	.0220	.0650	.0215			
2	.0375	.0455	.0375			
3	.0425	.0430	.0440			
4	.0380	.0380	.0470			
5	.0445	.0420	.0525			
10	.0505	.0340	.0525			
20	.0400	.0280	.0420			
30	.0480	.0450	.0540			
40	.0410	.0320	.0420			
50	.0420	.0390	.0460			

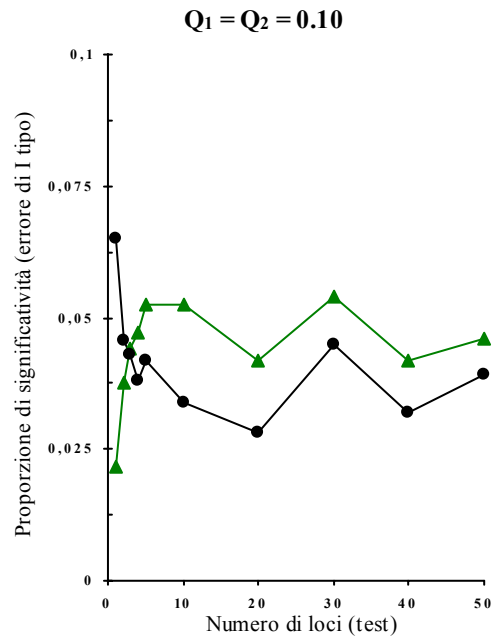
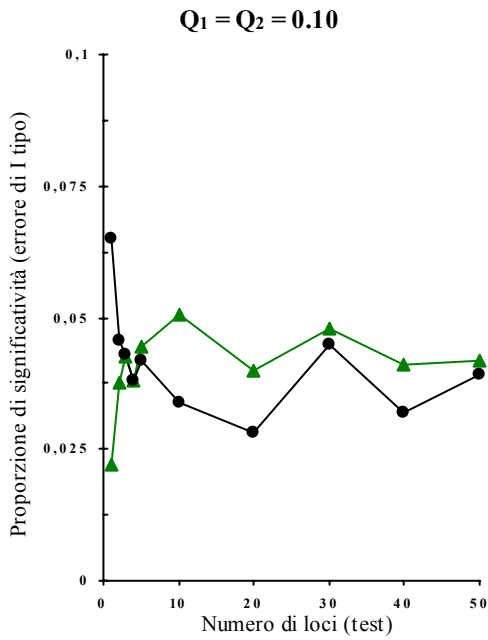
\* per k > 20, MC = 1000, B = 500



● :  $\sum_{h=1}^k X_h^2 \sim \chi_k^2$   
 ▲ :  $T_F^m = \psi(\lambda_T'', \lambda_F'', \lambda_L'')$   
**Figura 7.4a:** alternativa bidirezionale



● :  $\sum_{h=1}^k X_h^2 \sim \chi_k^2$   
 ▲ :  $T_F^m = \psi(\lambda_T^p, \lambda_F^p, \lambda_L^p)$   
**Figura 7.4b:** alternativa unidirezionale



### 7.7 STUDIO VIA SIMULAZIONE QUANDO LE DIFFERENZE RIGUARDANO $Z \leq k$ LOCI

Come anticipato nel paragrafo 7.5.2, è a priori non noto se vi sono differenze per quanto riguarda uno, alcuni o quasi tutti i loci (o nessuno nel caso in cui l'ipotesi nulla congiunta è vera). Pertanto

$$\forall k \in \mathbb{N}_+, \Pr\{Y_{hj} = A_h\} = \Pr\{X_{hj} = 1\} = Q_{hj}; \quad h = 1, \dots, k, j = 1, 2.$$

Per ciascuna popolazione  $j$ -esima ( $j = 1, 2$ ) verranno quindi generate  $2 \cdot n_j$  ( $n_j$  individui diploidi) prove bernoulliane con probabilità di successo pari a  $Q_{hj}$  per ogni locus  $h$ -esimo considerato ( $h = 1, \dots, k$ ). Si considererà un numero progressivamente crescente di loci  $z \leq k$  *sub* le relative sotto-ipotesi alternative  $H_{1h}$  ( $h = 1, \dots, z$ ), mentre i restanti  $(k - z)$  loci verranno considerati sotto  $H_0$ .

Nel presente studio, il numero dei loci  $k = 12$  (come nell'esemplificazione del paragrafo 7.3), le frequenze del primo campione  $Q_{h1} = Q_1 = 0.1$ , le  $(k - z)$  frequenze del secondo campione sotto  $H_0$   $Q_{m2} = Q_1 = 0.1$  ( $m = z + 1, \dots, k$ ) mentre le  $z$  frequenze  $Q_{h2}$  ( $h = 1, \dots, z$ ) considerate sotto  $H_1$  saranno uguali ( $Q_{h2} = Q_2$ ) per tutti i  $z$  loci ma varieranno al variare di  $z$  nel seguente modo (tabella 7.10):

**Tabella 7.11:** Frequenze dell'allele nei 12 loci

$z$	$k - z$	$Q_1$	$Q_2$
0	12	0.1	0.1
1	11	0.1	0.3
2	10	0.1	0.26
3	9	0.1	0.22
4, 5	8, 7	0.1	0.20
6, 7, 8	6, 5, 4	0.1	0.18
9, 10, 11, 12	3, 2, 1, 0	0.1	0.16

Le seguenti simulazioni Monte Carlo saranno effettuate con un livello di significatività  $\alpha = 0.05$ , un numero di simulazioni Monte Carlo  $MC = 2000$  e un numero di permutazioni (ricampionamenti condizionati CMC)  $B = 1000$ . La numerosità campionaria sarà uguale per i due campioni ( $n_1 = n_2$ ) e sarà pari a 50 (100 geni).

Per ogni procedura di verifica, nella tabella 7.11 vengono indicate le proporzioni di ripetizioni ottenute che rifiutano l'ipotesi nulla globale  $H_0$  per i 12 loci considerati; quando  $Q_1$  è differente da  $Q_2$  (almeno una sotto-ipotesi alternativa  $H_{1h}$  è vera) la

proporzione di rifiuti di  $H_0$  stima la potenza del test, mentre quando  $Q_1 = Q_2$  ( $H_0$  vera) stima l'errore di I tipo realizzato.

I test parziali e globali considerati sono gli stessi del paragrafo 7.6; anche in questo caso si presenteranno i risultati sia per alternative bidirezionali che per alternative unidirezionali.

I risultati sono presentati nelle tabelle 7.12 (caso unidirezionale) e 7.13 (caso bidirezionale): si può osservare che la soluzione iterata è prossima al *summation method* nel caso bidirezionale, mentre è preferibile nel caso unidirezionale.

*Osservazione 7.6:* In entrambi gli studi di simulazione presentati, vengono prese in considerazione alternative unidirezionali tali che  $H_1 : \{\cup_{h=1}^k [X_{h1} \stackrel{d}{<} X_{h2}]\}$ . L'unidirezionalità su tutti i loci è dovuta ad esigenze semplificatrici dell'impostazione dello studio di simulazione; nelle situazioni reali, se *a priori* si hanno informazioni sulla direzione della deviazione da  $H_0$  per ogni locus considerato, si strutturerà il problema in base a queste informazioni. Grazie alla flessibilità del metodo, è possibile considerare il caso di sotto-alternative bidirezionali e unidirezionali (in entrambe le direzioni) nel medesimo insieme di sotto-ipotesi.



**Tabella 7.12:** Studio di simulazione “→”;  $Q_1 = 0.10$ ;  $n_1 = n_2 = 50$ ;  $\alpha = 0.05$ ;  $MC = 2000$ ,  $B = 1000$ ;  $k = 12$ ;

$n^\circ$ loci sotto $H_1$ (●)	$Q_2$	$\psi_T(\lambda_1^p, \dots, \lambda_k^p)$	$\psi_F(\lambda_1^p, \dots, \lambda_k^p)$	$\psi_L(\lambda_1^p, \dots, \lambda_k^p)$	$\psi_T(\lambda_T^p, \lambda_F^p, \lambda_L^p)$	$\sum_{h=1}^k X_h^2$
$H_0$ VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)						
○○○○○○○○○○○○	0.10	0.0530	0.0565	0.0505	0.0550	0.0430
$H_0$ FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)						
●○○○○○○○○○○	0.30	0.8685	0.4420	0.2020	0.8160	0.6580
●●○○○○○○○○	0.26	0.8765	0.8080	0.4610	0.8610	0.8165
●●●○○○○○○	0.22	0.7650	0.8495	0.6275	0.8305	0.7680
●●●●○○○○○○	0.20	0.7150	0.8885	0.7325	0.8530	0.7695
●●●●●○○○○	0.20	0.7815	0.9575	0.8900	0.9385	0.8665
●●●●●●○○○○	0.18	0.6540	0.9455	0.8730	0.9075	0.7755
●●●●●●●○○○	0.18	0.7010	0.9780	0.9480	0.9610	0.8485
●●●●●●●●○○	0.18	0.7445	0.9885	0.9820	0.9835	0.8980
●●●●●●●●●○○	0.16	0.5415	0.9460	0.9480	0.9290	0.7015
●●●●●●●●●●○	0.16	0.5830	0.9770	0.9765	0.9725	0.7630
●●●●●●●●●●●○	0.16	0.6215	0.9820	0.9900	0.9800	0.8050
●●●●●●●●●●●●	0.16	0.6490	0.9945	0.9980	0.9935	0.8440

**Tabella 7.13:** Studio di simulazione “↔”;  $Q_1 = 0.10$ ;  $n_1 = n_2 = 50$ ;  $\alpha = 0.05$ ;  $MC = 2000$ ,  $B = 1000$ ;  $k = 12$ ;

$n^\circ$ loci sotto $H_1$ (●)	$Q_2$	$\psi_T(\lambda_1, \dots, \lambda_k)$	$\psi_F(\lambda_1, \dots, \lambda_k)$	$\psi_L(\lambda_1, \dots, \lambda_k)$	$\psi_T(\lambda_T^n, \lambda_F^n, \lambda_L^n)$	$\sum_{h=1}^k X_h^2$
$H_0$ VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)						
○○○○○○○○○○○○	0.10	0.0475	0.0445	0.0490	0.0485	0.0505
$H_0$ FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)						
●○○○○○○○○○○	0.30	0.8130	0.4220	0.2125	0.7570	0.6580
●●○○○○○○○○	0.26	0.8140	0.7370	0.4455	0.8005	0.8165
●●●○○○○○○	0.22	0.6600	0.7370	0.5395	0.7170	0.7680
●●●●○○○○○○	0.20	0.5825	0.7740	0.6075	0.7105	0.7695
●●●●●○○○○	0.20	0.6690	0.8615	0.7620	0.8250	0.8665
●●●●●●○○○○	0.18	0.5100	0.7650	0.6795	0.7240	0.7755
●●●●●●●○○○	0.18	0.5540	0.8470	0.7910	0.8095	0.8485
●●●●●●●●○○	0.18	0.5965	0.9025	0.8645	0.8590	0.8980
●●●●●●●●●○○	0.16	0.4045	0.7085	0.6595	0.6635	0.7015
●●●●●●●●●●○	0.16	0.4110	0.7705	0.7275	0.7125	0.7630
●●●●●●●●●●●○	0.16	0.4680	0.8175	0.7940	0.7770	0.8050
●●●●●●●●●●●●	0.16	0.4730	0.8610	0.8345	0.8090	0.8440

## 7.8 LEGGE DI HARDY-WEINBERG

La legge di Hardy-Weinberg rappresenta una situazione ideale che probabilmente non esiste in natura. Essa dimostra che, nelle grosse popolazioni, il processo di eredità non causa cambiamenti nelle frequenze alleliche che restano costanti da una generazione all'altra.

Useremo l'esempio dei moscerini della frutta (*Drosophila melanogaster*) per spiegare la legge di Hardy-Weinberg. Un semplice incrocio tra moscerini neri e grigi dimostra che l'allele  $B$  per il colore grigio del corpo, è dominante sull'allele  $b$  per il colore nero del corpo. I moscerini con il corpo grigio ne comprendono alcuni che sono omozigoti  $BB$  ed alcuni che sono eterozigoti  $Bb$ . Ovviamente, tutti i moscerini neri sono omozigoti  $bb$ . Poiché per il gene in questione, indichiamo con  $p$  la frequenza dell'allele  $B$ , e con  $q$  la frequenza dell'allele  $b$  nella popolazione, otterremo un'equazione binomiale,  $p + q = 1$ , quindi  $(p + q)^2 = 1$ . Questa equazione può essere sviluppata per descrivere il rapporto fra le frequenze alleliche e i genotipi della popolazione:

$$\begin{array}{ccccccc}
 p^2 & + & 2pq & + & q^2 & = & 1 \\
 \text{frequenza} & & \text{frequenza} & & \text{frequenza} & & \text{tutti gli individui} \\
 \text{di } BB & & \text{di } Bb & & \text{di } bb & & \text{in una popolazione}
 \end{array}$$

Ogni popolazione in cui la distribuzione genotipica coincida alla relazione  $p^2 + 2pq + q^2 = 1$  è in equilibrio genetico. Il principio dell'equilibrio genetico ci dice cosa ci si deve attendere quando una popolazione che si riproduce sessualmente non è in evoluzione. Le proporzioni degli alleli saranno sempre le stesse a patto che si riscontrino le seguenti condizioni:

1. accoppiamento casuale
2. assenza di mutazioni
3. popolazioni infinitivamente grandi
4. assenza di migrazioni
5. assenza di selezione

L'assunto che gli alleli all'interno degli individui sono indipendenti è quindi giustificato dall'equilibrio di Hardy-Weinberg, relativo a popolazioni virtuali con un numero di individui abbastanza grande. E' quindi giustificato anche per popolazioni reali, generalmente abbastanza piccole, soggette a un complesso processo di selezione? Certamente, sotto l'assunto di indipendenza, molti problemi probabilistici associati sono più semplici da risolvere rispetto al caso di dipendenza. Quindi bisogna tenere in considerazione che nei contesti reali l'assunto di indipendenza è approssimativamente valido, conducendo a soluzioni approssimate.

## 7.9 STUDIO VIA SIMULAZIONE NEL CASO DI DIPENDENZA FRA I $k$ LOCI

Nelle simulazioni precedenti si è ipotizzata l'indipendenza fra i  $k$  loci, requisito utile per il *summation method* in quanto richiede l'indipendenza delle statistiche componenti, ma irrilevante per la metodologia di combinazione non parametrica.

Si svolgerà quindi un'analisi nel caso di dipendenza fra i  $k$  loci secondo il seguente modello :

$$X_{hji} \in \{0, 1\} \quad i = 1, \dots, 2 \cdot n_j, j = 1, 2, h = 1, \dots, k$$

$$X_{hji} = \mathbf{I}\{(-0.8 + \rho \cdot X_{(h-1)ji} + U(-1, 1) \geq 0\} = \begin{cases} 1 & \text{se } (-0.8 + \rho \cdot X_{(h-1)ji} + U(-1, 1) \geq 0 \\ 0 & \text{se } (-0.8 + \rho \cdot X_{(h-1)ji} + U(-1, 1) < 0 \end{cases}$$

$$0 \leq \rho \leq 1.8, \quad i = 1, \dots, 2 \cdot n_j, j = 1, 2, h = 1, \dots, k$$

$$\text{e } \Pr\{X_{0ji} = 1 \mid H_1\} = \phi \quad \text{con } 0.10 \leq \phi < 1, \quad i = 1, \dots, 2 \cdot n_j, j = 1, 2$$

Chiaramente  $X_{0ji}$  non rappresenta alcun locus, ma semplicemente una prova aleatoria.

$$\Pr\{\rho = 0 \mid H_0\} = 1 \Rightarrow \Pr\{X_{hji} = 1 \mid H_0\} = 0.1 \quad i = 1, \dots, 2 \cdot n_j, j = 1, 2, h = 1, \dots, k$$

$$\begin{aligned} \Pr\{X_{1ji} = 1 \mid H_1\} &= \Pr\{X_{1ji} = 1 \mid X_{0ji} = 0 \cap H_1\} \Pr\{X_{0ji} = 0 \mid H_1\} + \\ &\quad \Pr\{X_{1ji} = 1 \mid X_{0ji} = 1 \cap H_1\} \Pr\{X_{0ji} = 1 \mid H_1\} \\ &= 0.1 \times (1 - \phi) + (0.1 + \phi/2) \times \phi = 0.1 + (\rho/2 \times \phi) \end{aligned}$$

$$\text{Affinchè } \Pr\{X_{1ji} = 1 \mid H_1\} = \phi \Rightarrow \rho = 2 - (0.2/\phi) \Rightarrow \Pr\{X_{hji} = 1 \mid H_1\} = \phi$$

$$i = 1, \dots, 2 \cdot n_j, j = 1, 2, h = 1, \dots, k$$

$$\Pr\{X_{hji} = 1 \mid X_{(h-1)ji} = 1 \cap H_1\} = 0.1 + \rho/2 = 1.1 - (0.1/\phi)$$

$$i = 1, \dots, 2 \cdot n_j, j = 1, 2, h = 1, \dots, k$$

$$\Pr\{X_{hji} = 1 \mid X_{(h-1)ji} = 0 \cap H_1\} = 0.1$$

$$i = 1, \dots, 2 \cdot n_j, j = 1, 2, h = 1, \dots, k$$

Riassumiamo ciò nella tabella sottostante (tabella 7.14):

**Tabella 7.14:** Modello di dipendenza dei  $k$  loci sotto  $H_1$

<i>prova aleatoria</i>	<i>locus 1</i>	<i>locus h, h = 2, ..., k</i>
$\left. \begin{array}{l} \Pr\{X_{0ji}=\bullet\}=\phi \\ \Pr\{X_{0ji}=\circ\}=(1-\phi) \end{array} \right\} \text{ ?}$	$\left\{ \begin{array}{l} \Pr\{X_{1ji}=\bullet\} \\ \Pr\{X_{1ji}=\circ\} \end{array} \right\}$	$\left\{ \begin{array}{l} \Pr\{X_{hji}=\bullet\} \\ \Pr\{X_{hji}=\circ\} \end{array} \right\}$
	$\left\{ \begin{array}{l} 1.1-(0.1/\phi) \text{ se } X_{0ji}=\bullet \\ 0.10 \text{ se } X_{0ji}=\circ \\ (0.1/\phi)-0.1 \text{ se } X_{0ji}=\bullet \\ 0.90 \text{ se } X_{0ji}=\circ \end{array} \right\}$	$\left\{ \begin{array}{l} 1.1-(0.1/\phi) \text{ se } X_{(h-1)ji}=\bullet \\ 0.10 \text{ se } X_{(h-1)ji}=\circ \\ (0.1/\phi)-0.1 \text{ se } X_{(h-1)ji}=\bullet \\ 0.90 \text{ se } X_{(h-1)ji}=\circ \end{array} \right\}$

Si può osservare che, sotto  $H_1$ , per l'unità  $i$ -esima del  $j$ -esimo gruppo la probabilità di ottenere un'allele raro ( $\bullet$ ) nel locus  $h$ -esimo sarà la stessa che si verifica sotto  $H_0$

( $\Pr\{X_{hji} = 1\} = 0.1$ ) se nel locus  $(h-1)$ -esimo si è ottenuto un allele comune ( $X_{hji} = 0 = \circ$ ).

Diversamente, se nel locus  $(h-1)$ -esimo si è ottenuto un allele raro,  $\Pr\{X_{hji} = 1 | X_{(h-1)ji} = 1 \cap H_1\} = 1.1 - (0.1/\phi)$ .

Con riferimento agli studi di simulazione dei precedenti paragrafi, per  $\phi$  pari a 0.15, 0.16, 0.18, 0.20, 0.22, 0.26 corrispondono rispettivamente  $\Pr\{X_{hji} = 1 | X_{(h-1)ji} = 1 \cap H_1\} \cong 0.43, 0.48, 0.54, 0.60, 0.65, 0.72$ .

Si riprodurrà pertanto una sintesi degli studi via simulazione dei paragrafi 7.6 e 7.7 escludendo ovviamente il caso  $k = 1$  e generando i dati secondo il modello sopra descritto.

La sintesi comprende:

- analogamente allo studio via simulazione del paragrafo 7.6 con  $k = 2, 5, 10, 20$   $Q_1 = 0.10, E(Q_2) = 0.15$ , alternative bidirezionali, considerando il test globale  $T_F''' = \psi_F(\lambda_T'', \lambda_F'', \lambda_L'')$ ;
- analogamente allo studio via simulazione del paragrafo 7.7 con  $k = 12$  e  $z = 3, 5, 7, 10$ ,  $Q_1 = 0.10$ ,  $E(Q_2) = 0.22, 0.20, 0.18, 0.16$  rispettivamente ai valori di  $z$ , alternative bidirezionali, considerando il test globale  $T_F''' = \psi_F(\lambda_T'', \lambda_F'', \lambda_L'')$ ;
- si valuterà inoltre, sotto  $H_0$ , la stima dell'errore di I tipo, con  $E(Q_1) = E(Q_2) = 0.15, 0.18, 0.20, 0.22, 0.26$  e con la struttura di dipendenza sopra descritta;

**Tabella 7.15:** studio via simulazione nel caso di dipendenza fra i  $k$  loci

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)					
$Q_1 = 0.10, E(Q_2) = 0.15; n_1 = n_2 = 20$ (40 geni); $\alpha = 0.05; MC = 1000, B = 1000; k = 2, 5, 10, 20$					
loci	$\psi_F(\lambda_T^{\prime\prime}, \lambda_F^{\prime\prime}, \lambda_L^{\prime\prime})$	$\sum_{h=1}^k X_h^2$	$\psi_T(\lambda_1, \dots, \lambda_k)$	$\psi_F(\lambda_1, \dots, \lambda_k)$	$\psi_L(\lambda_1, \dots, \lambda_k)$
2	.0900	.1140	.0650	.0880	.0860
5	.1650	.1740	.1130	.1720	.1440
10	.2080	.2230	.1230	.2180	.1840
20	.2960	.3130	.1290	.3030	.2430

**Tabella 7.16:**  $Q_1 = 0.10, E(Q_2) = .26, .22, .20, .18, .16; n_1 = n_2 = 50$  (100 geni);  $\alpha = 0.05; MC = 1500, B = 750;$

$n^\circ$ loci sotto H <sub>1</sub>	$\psi_T(\lambda_1, \dots, \lambda_k)$	$\psi_F(\lambda_1, \dots, \lambda_k)$	$\psi_L(\lambda_1, \dots, \lambda_k)$	$\psi_F(\lambda_T^{\prime\prime}, \lambda_F^{\prime\prime}, \lambda_L^{\prime\prime})$	$\sum_{h=1}^k X_h^2$
H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)					
oooooooooooo	.0550	.0610	.0450	.0530	.0570
H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)					
●●●oooooooooo	.6110	.7100	.3410	.6940	.7360
●●●●ooooooooo	.5950	.7980	.5330	.7690	.8090
●●●●●oooooo	.5210	.7590	.5990	.7440	.7700
●●●●●●ooooo	.3690	.7030	.5890	.6800	.7130

**Tabella 7.17:**  $E(Q_1) = E(Q_2) = .15, .18, .20, .26, .30; n_1 = n_2 = 50$  (100 geni);  $\alpha = 0.05; MC = 1000, B = 1000;$

$n^\circ$ loci sotto H <sub>1</sub>	$\psi_T(\lambda_1, \dots, \lambda_k)$	$\psi_F(\lambda_1, \dots, \lambda_k)$	$\psi_L(\lambda_1, \dots, \lambda_k)$	$\psi_F(\lambda_T^{\prime\prime}, \lambda_F^{\prime\prime}, \lambda_L^{\prime\prime})$	$\sum_{h=1}^k X_h^2$
H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)					
oooooooooooo	.0520	.0530	.0490	.0500	.0760
oooooooooooo	.0420	.0530	.0600	.0540	.0770
oooooooooooo	.0460	.0510	.0540	.0560	.0900
oooooooooooo	.0470	.0510	.0490	.0530	.1080
oooooooooooo	.0320	.0450	.0490	.0440	.1070

Si può osservare dalla tabella 7.17 che il test  $\sum_{h=1}^k X_h^2$  risulti anticonservativo, a causa della mancata assunzione di indipendenza fra i loci: risulta infatti  $k \cdot \chi_{\alpha,1}^2 \geq \sum_{h=1}^k X_h^2 \geq \chi_{\alpha,k}^2$  (con le uguaglianze verificate rispettivamente nel caso di dipendenza completa e indipendenza); verrebbe così spiegata la maggiore potenza ottenuta nei casi presentati nelle due tabelle precedenti (tabella 7.15 e 7.16).

## 7.10 CONTROLLO DELLA MOLTEPLICITÀ

Se l'ipotesi nulla globale di omogeneità delle frequenze degli alleli nelle due popolazioni viene rifiutata, probabilmente sarà di interesse la valutazione delle inferenze marginali, e cioè *quali* loci presentano un'eterogeneità significativa tra le due popolazioni. Come anticipato nell'osservazione 7.3, ciò non può essere fatto semplicemente esaminando i  $p$ -values parziali, ma si necessita di una loro "correzione". L'obiettivo delle procedure di *multiple testing* è di controllare il "*Familywise Error Rate*", che è la probabilità massima che una o più ipotesi nulle vengano erroneamente rifiutate.

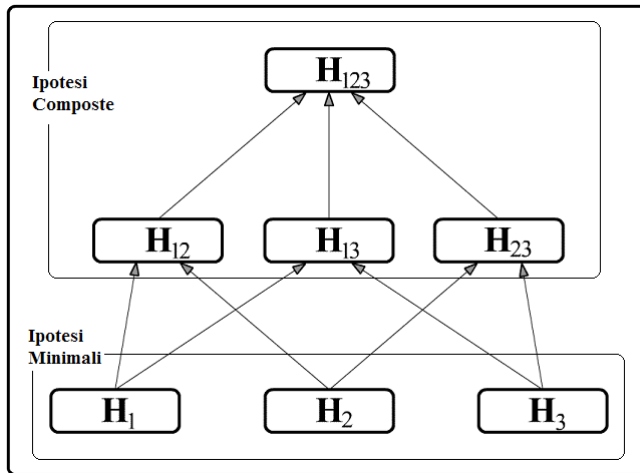
Negli anni recenti l'area del *multiple testing* è stata apparentemente pilotata dalle applicazioni di "*closed testing*". Questi metodi sono del tipo "*step-wise*", e di solito sono più potenti delle loro controparti "*single-step*". Un esempio di un metodo "*single-step*" è il metodo semplice di Bonferroni, in cui le inferenze multiple devono utilizzare livelli di significatività  $\alpha/k$ , dove  $k$  indica il numero di inferenze distinte e dove  $\alpha$  indica il tasso massimo di errore di I tipo permesso all'insieme delle  $k$  inferenze. Un metodo "*step-wise*" di solito utilizza livelli critici più grandi di  $\alpha/k$  permettendo più spesso di rilevare differenze significative, e quindi comporta una maggiore potenza.

### 7.10.1 METODI DI *CLOSED TESTING* PER TEST MULTIPLI

L'idea dietro il *closed testing* è semplice, ma richiede qualche notazione.

Si supponga di volere verificare le ipotesi  $H_1, H_2$  e  $H_3$ . Il metodo funziona come segue:

1. Si verifichi ogni ipotesi,  $H_1, H_2, H_3$  utilizzando un test di livello  $\alpha$  appropriato;
2. Creare la "chiusura" dell'insieme, che è l'insieme di tutte le possibili intersezioni tra  $H_1, H_2, H_3$ , in questo caso le ipotesi  $H_{12}, H_{13}, H_{23}$  e  $H_{123}$ ;
3. Si verifichi ogni intersezione utilizzando un test di livello  $\alpha$  appropriato;
4. È possibile rifiutare qualsiasi ipotesi  $H_i$ , con controllo del FWE, quando entrambe le seguenti condizioni reggono:
  - Il test per  $H_i$ , stesso produce un risultato statisticamente significativo;
  - Il test per ogni ipotesi di intersezione che include  $H_i$ , è statisticamente significativo;



**Figura 7.5:** Albero delle ipotesi (Westfall e Wolfinger, 2000)

Utilizzando una procedura di *closed testing*, il  $p$ -value aggiustato per un'ipotesi data  $H_i$ , è il massimo fra tutti i  $p$ -values per i test che includono  $H_i$ , come caso speciale ( includendo il  $p$ -value per il medesimo test  $H_i$  ).

$\Rightarrow$   $p$ -value aggiustato =  $\max(p_j, p_{jk}, p_{jkl}, \dots)$  cioè il massimo tra tutte le ipotesi in cui  $j$  è componente.

Una procedura *step-down* prende in considerazione prima l'ipotesi globale (intersezione di tutte le ipotesi minimali) per poi procedere a ritroso secondo tutte le ipotesi implicate. Se un'ipotesi non viene rigettata, si accetteranno tutte le ipotesi (minimali e non) che da essa sono implicate senza ulteriori test; quindi una ipotesi viene valutata se e solo se tutte le ipotesi che la includono sono rigettate. Si presenteranno brevemente nel seguito gli *shortcut* per procedure di *closed testing* per la Procedura *step-down* MinP di Bonferroni-Holm e quella *step-down* di Tippett (Finos *et al.*, 2000).

#### 7.10.2 SHORTCUT DELLA PROCEDURA STEP-DOWN MIN-P BONFERRONI-HOLM

Questa è senza dubbio la più classica delle procedure a più livelli, è detta anche "*Sequentially Rejective Bonferroni procedure*", proposta da Holm (1979) come semplice applicazione del metodo di Bonferroni. Holm dimostra che non è necessario calcolare i  $p$ -values per l'intero albero: è necessario calcolare i  $p$ -values per i nodi dell'albero corrispondenti ai  $p$ -values ordinati.

$p_{(1)}, \dots, p_{(k)}$  è il vettore di  $p$ -value ordinati (crescenti) delle ipotesi minimali.

1.  $adj\_p_{(1)} = k \cdot p_{(1)}$ ; se  $adj\_p_{(1)} \leq \alpha$  si rifiuta  $H_{(1)}$  e si procede; altrimenti si accetta  $H_{(1)}, \dots, H_{(k)}$  e STOP;

2.  $adj\_p_{(i)} = \max((k - i + 1) \cdot p_{(i)}, adj\_p_{(i-1)})$ ; se  $adj\_p_{(i)} \leq \alpha$  si rifiuta (anche)  $H_{(i)}$  e si procede; altrimenti accetta  $H_{(i)}, \dots, H_{(k)}$  e STOP;  $i = 2, \dots, k$

**Figura 7.6:** procedura *step-down* MinP Bonferroni-Holm (Finos *et al.*, 2000)

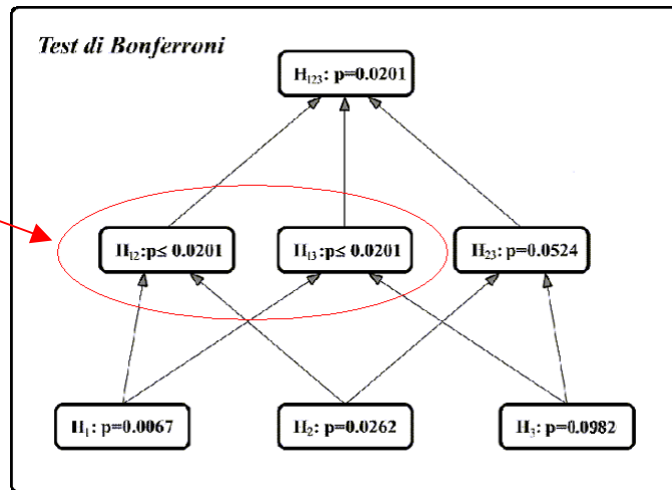
Non serve testarle

Infatti valgono:

$$p_{123} \geq p_{12} \geq p_1 \text{ (infatti } 3 \cdot p_1 \geq 2 \cdot p_1 \geq p_1)$$

$$p_{123} \geq p_{13} \geq p_1 \text{ (infatti } 3 \cdot p_1 \geq 2 \cdot p_1 \geq p_1)$$

$$p_{23} \geq p_2 \text{ (infatti } 2 \cdot p_2 \geq p_2)$$



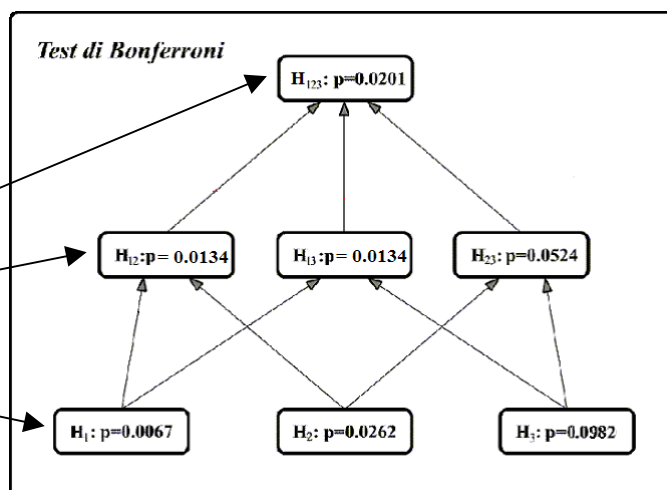
Nell'esempio a cui ci si riferisce, il vettore ordinato delle significatività delle ipotesi minimali risulta  $p_1 < p_2 < p_3$ ; si procederà poi all'ipotesi globale  $H_{123} : p_{H_{123}} = 3 \cdot p_3 = 3 \cdot 0.0067 = 0.0201$ . Ponendo  $\alpha = 0.05$ , si rifiuterà  $H_{123}$  ed automaticamente tutte le ipotesi di cui  $H_3$  è componente, e cioè  $H_3, H_{13}, H_{23}$ : all'ipotesi  $H_3$  si assocerà quindi una significatività di 0.0201. Il passo successivo consisterà nel valutare l'intersezione delle ipotesi non (ancora) rigettate e cioè  $H_{12}$ , la cui significatività è pari a  $p_{H_{12}} = 2 \cdot 0.0262 = 0.0524$ . Essendo  $\alpha = 0.05$  le due ipotesi  $H_1$  e  $H_2$  vengono accettate.

**Figura 7.7:** procedura di Bonferroni *single-step* (Finos *et al.*, 2000)

$$p_{123} = 3 \cdot \min(p_1, p_2, p_3)$$

$$p_{12} = 2 \cdot \min(p_i, p_j)$$

t di Student



Se avessimo applicato la disuguaglianza di Bonferroni (figura 7.x) ad una procedura *single-step* ( $p_{H_i} : \max(p_{H_{bj}} : H_j \supseteq H_i)$ ), avremmo ottenuto gli stessi risultati ma si sarebbero calcolate le significatività di tutte le ipotesi invece di due sole.



7.10.3 *SHORTCUT* DELLA PROCEDURA *STEP-DOWN* DI TIPPETT

Il metodo di Bonferroni Holm risulta essere troppo conservativo quando la struttura di correlazione fra le variabili è molto forte. Invece di confrontare l'osservato  $\min p$  di una data ipotesi composta con  $\alpha/k$  ( $k$  sta per il numero di ipotesi minimali) Westfall e Young (1993) suggeriscono di confrontarlo con l' $\alpha$ -quantile della distribuzione MinP sotto l'ipotesi nulla. Formalmente, questo equivale a calcolare il  $p$ -value  $p = \Pr(\text{MinP} \leq \min p)$ , dove MinP denota la variabile casuale del minimo  $p$ -value dell'ipotesi composta, e  $\min p$  indica la significatività minima osservata per l'ipotesi in considerazione.

Normalmente, la distribuzione del MinP è sconosciuta ma può essere facilmente approssimata con metodi *bootstrap*, come mostrato da Westfall e Young, o con metodi di ricampionamento senza ripetizione, come mostrato in Finos *et al.* (2000).

Westfall e Young inoltre mostrano che, parimenti alla procedura di Bonferroni-Holm, questo metodo può essere utilizzato per generare una procedura *step-wise*. Non sarà quindi necessario calcolare tutti i  $p$ -value, semplicemente basterà utilizzare l'algoritmo di sezione parallelo a quello di Holm.

Un'ultima nota: la stima dell' $\alpha$ -quantile del MinP via ricampionamento senza ripetizione equivale alla combinazione non parametrica di Tippett: la significatività viene calcolata come

$p = \Pr(\min_{1 \leq i \leq k}(\lambda_i)^* \leq \min_{1 \leq i \leq k}(\lambda_i))$  dove per  $\min_{1 \leq i \leq k}(\lambda_i)^*$  intendiamo la distribuzione di MinP e con  $\min_{1 \leq i \leq k}(\lambda_i)$  il minimo dei  $p$ -value dell'ipotesi composta ( $\min p$ ).

$$p = \Pr(\min_{1 \leq i \leq k}(\lambda_i)^* \leq \min_{1 \leq i \leq k}(\lambda_i)) = \Pr(1 - \min_{1 \leq i \leq k}(\lambda_i)^* \geq 1 - \min_{1 \leq i \leq k}(\lambda_i)) = \Pr(\max_{1 \leq i \leq k}(\lambda_i)^* \leq \max_{1 \leq i \leq k}(\lambda_i)) = T_T''$$

dove  $T_T''$  indica appunto la combinazione di Tippett.

La *shortcut* è la seguente:

$p_{(1)}, \dots, p_{(k)}$  è il vettore di  $p$ -value ordinati (crescenti) delle ipotesi minimali.

1.  $adj\_p_{(1)} = p_{(1)\dots(k)Tippett}''$ ; se  $adj\_p_{(1)} \leq \alpha$  si rifiuta  $H_{(1)}$  e si procede; altrimenti si accetta  $H_{(1)}, \dots, H_{(k)}$  e STOP;
2.  $adj\_p_{(i)} = \max(p_{(i)\dots(k)Tippett}'', adj\_p_{(i-1)})$ ; se  $adj\_p_{(i)} \leq \alpha$  si rifiuta (anche)  $H_{(i)}$  e si procede; altrimenti accetta  $H_{(i)}, \dots, H_{(k)}$  e STOP;  $i = 2, \dots, k$

# Capitolo 8

## Verifica dell'uguaglianza di due distribuzioni: un'applicazione della procedura multi-aspect

### 8.1 INTRODUZIONE

In generale, confrontare due cose significa verificare se esse sono uguali o diverse e, in quest'ultimo caso, misurare quanto è la diversità. Per confrontare dunque due distribuzioni, secondo uno stesso carattere ma relative a diversi collettivi, è necessario stabilire un criterio di uguaglianza fra le distribuzioni.

Il confronto può essere effettuato direttamente sulle due distribuzioni, senza cioè ricorrere preventivamente a delle loro sintesi. Ci si può aspettare, però, che il confronto possa essere eseguito efficacemente ricorrendo ad indici con cui si sintetizzano le due distribuzioni: ad esempio, due distribuzioni possono essere uguali rispetto alla media, o più in generale, uguali rispetto ad un parametro di posizione.

Il problema può quindi essere affrontato in diversi modi, in accordo con le specifiche *side-assumptions* riguardanti il significato del concetto di ineguaglianza in distribuzione. Si presenteranno nei paragrafi 8.2.1 e 8.2.2 particolari specificazioni di questo concetto.

### 8.2 VERIFICA DELL'UGUAGLIANZA DI DUE DISTRIBUZIONI

Si assuma che le osservazioni di una variabile risposta  $X$  su  $n$  unità siano suddivise in due gruppi corrispondenti a due livelli di un trattamento simbolico, di  $n_1$  ed  $n_2$  unità rispettivamente. Si assuma inoltre che le variabili risposta nei due gruppi seguano distribuzioni non note  $P_1$  e  $P_2$ , entrambe definite nello spazio probabilistico  $(\Omega, \beta)$ , dove  $\Omega$  rappresenta lo spazio campionario e  $\beta$  un'algebra di eventi. Quindi, i dati campionari sono

$$\mathbf{X} = \{X_{ji}; i = 1, \dots, n_j, j = 1, 2\}$$

In generale è di interesse verificare l'ipotesi nulla  $H_0 : \{P_1 \stackrel{d}{=} P_2\}$ , che afferma che i due gruppi hanno la medesima distribuzione sottostante, contro l'alternativa

$H_1 : \{P_1 <^d P_2\}$ . D'ora in avanti, senza perdita di generalità, si consideri l'alternativa unidirezionale  $H_1 : \{P_1 <^d P_2\}$  tale che sia soddisfatta la condizione di dominanza stocastica di  $X_2$  rispetto a  $X_1$  per la quale le due funzioni di ripartizione  $F_1$  e  $F_2$  sono tali che,  $\forall x \in \mathfrak{R}^1, F_1(x) \geq F_2(x)$ . Si assuma inoltre che le due distribuzioni sottostanti siano continue, in conformità ai problemi del tipo *goodness of fit*.

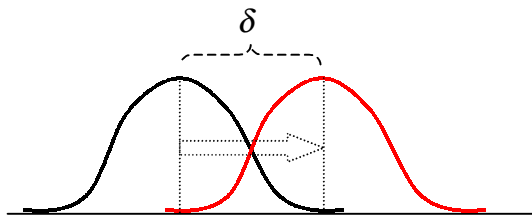
### 8.2.1 MODELLO DI RISPOSTA OMOSCHEDASTICO CON EFFETTI ADDITIVI FISSI

La prima specificazione del concetto di ineguaglianza in distribuzione si riferisce al cosiddetto confronto di due medie (o, più in generale, di due posizioni<sup>1</sup>). Le *side-assumptions* per il problema sono tali che le variabili risposta si comportano secondo il seguente modello additivo:

$$(M.i.o) \quad X_{ji} = \mu + \delta_j + \sigma \cdot Z_{ji} \quad \begin{cases} X_{1i} = \mu + \sigma \cdot Z_{1i} \\ X_{2i} = \mu + \delta_2 + \sigma \cdot Z_{2i} \end{cases} \quad i = 1, \dots, n_j, j = 1, 2$$

dove  $\mu$  è una costante di disturbo della popolazione,  $Z_{ji}$  sono errori casuali scambiabili con distribuzione comune  $P$ ,  $\sigma$  è una costante di scala non dipendente né dalle unità né dai livelli del trattamento,  $\delta = \delta_2 (> 0)$  rappresenta l'effetto trattamento (o il cosiddetto funzionale di posizione<sup>2</sup> o [pseudo-] parametro di spostamento<sup>3</sup>). Si noti l'omoschedasticità delle risposte, che comporta  $F_1(x) = F_2(x - \delta_2), \forall x \in \mathfrak{R}^1$ , e quindi  $F_1(x) \geq F_2(x), \forall x \in \mathfrak{R}^1$ .

Si assuma perciò che sotto  $H_1$  la distribuzione di  $X_2$  risulti spostata di una quantità  $\delta > 0$  rispetto a quella di  $X_1$  (figura 8.1), così che  $H_1 : \{X_1 <^d (X_2 - \delta)\}$ . Con riferimento alle alternative unidirezionali, le ipotesi possono essere scritte in maniera equivalente come  $H_0 : \{\delta = 0\}$  e  $H_1 : \{\delta > 0\}$ .



**Figura 8.1:** confronto di due distribuzioni per il modello con effetti additivi fissi con omoschedasticità (M.i.o) sotto l'ipotesi alternativa  $H_1$ ;

<sup>1</sup> locations;

<sup>2</sup> location functional;

<sup>3</sup> shift [pseudo-] parameter;

## 8.1.2 MODELLO DI RISPOSTA CON EFFETTI STOCASTICI GENERALIZZATI

La seconda specificazione più generale per verificare  $H_0 : \{X_1 \stackrel{d}{=} X_2\}$  contro  $H_1 : \{X_1 \stackrel{d}{<} X_2\}$  può essere generalmente espressa secondo un modello con effetti stocastici generalizzati:

$$(M.ii) \quad X_{ji} = \mu + \Delta_{ji} + Z_{ji} \quad \begin{cases} X_{1i} = \mu + Z_{1i} \\ X_{2i} = \mu + \Delta_{2i} + Z_{2i} \end{cases} \quad i = 1, \dots, n_j, j = 1, 2$$

dove  $\mu$  è una costante di disturbo della popolazione,  $Z_{ji}$  sono errori casuali scambiabili con distribuzione comune  $P$  tali che  $\mu + Z_{ji} > 0$  in probabilità,  $\Delta_{ji}$  sono effetti stocastici generalizzati per il  $j$ -esimo livello di trattamento, che sono assunti essere uguali a zero per il primo gruppo ( $\Delta_{1i} = 0$ ) e non negativi con probabilità uno per il secondo gruppo ( $\Delta_{2i} \stackrel{P}{\geq} 0$ ), inoltre possono dipendere da  $(\mu, Z_{ji})$  e soddisfano all'ulteriore condizione  $(\mu + \Delta_{2i} + Z_{2i})^2 \geq (\mu + Z_{2i})^2, i = 1, \dots, n_2$ .

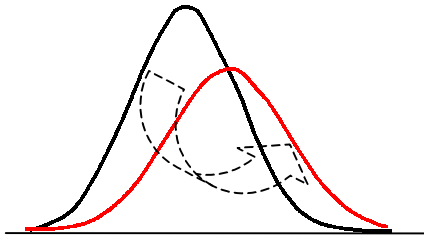
*Osservazione 1.8:* L'assunto di non negatività degli effetti stocastici  $\Delta_{ji}$  è equivalente alla condizione di dominanza delle due risposte. Quindi, le due funzioni di ripartizione  $F_1$  e  $F_2$  sono tali che,  $\forall x \in \mathfrak{R}^1, F_1(x) \geq F_2(x)$ .

Il modello (M.ii) è consistente con ogni tipo di dominanza stocastica. In particolare:

(M.ii) = (M.i.o) quando, con probabilità uno,  $\Delta_{2i} = \delta_2$ ;

$$(M.ii) \quad X_{ji} = \mu + \delta_i + W_{ji} + Z_{ji} \quad \begin{cases} X_{1i} = \mu + Z_{1i} \\ X_{2i} = \mu + \delta_2 + W_{2i} + Z_{2i} \end{cases} \quad i = 1, \dots, n_j, j = 1, 2$$

con  $\Delta_{2i} = \delta_2 + W_{2i} \geq 0$  dove  $\delta_2$  rappresenta l'effetto medio, e le quantità casuali (errori stocastici individuali sugli effetti)  $W$  possono dipendere dagli errori  $Z$ . Quando  $W$  e  $Z$  sono dipendenti, il trattamento può agire su più d'uno degli aspetti delle distribuzioni sottostanti, pertanto si è interessati ad estendere le analisi su più aspetti in conformità all'approccio multi-aspect (figura 8.2):



**Figura 8.2:** confronto di due distribuzioni per il modello con effetti stocastici generalizzati (M.ii) sotto l'ipotesi alternativa  $H_1$ ;

$X_{1i} = Z_{1i}$   
 $X_{2i} = \Delta_{2i} + Z_{2i} = \delta_2 + W_{2i} + Z_{2i} \quad i = 1, \dots, n_j, j = 1, 2$   
 $Z_{ji} = Z_{ji1} + Z_{ji2} + Z_{ji3}; Z_{jiz} \sim U(0, 8) \quad z = 1, \dots, 4$   
 $W_{2i} = a \cdot Z_{2i1} + b \cdot Z_{2i4}; a = 0.3, b = 0.25, \delta_2 = 1, 25$   
 $E[X_1] = 12, 1; \text{standard deviation} = 3, 96; n_1 = 1000$   
 $E[X_2] = 15, 6; \text{standard deviation} = 4, 40; n_2 = 1000$   
 Si veda il capitolo 8.11;

### 8.3 PROCEDURA MULTI-ASPECT PER LA VERIFICA DELL'UGUAGLIANZA DI DUE DISTRIBUZIONI

Le ipotesi nulle da verificare sono rispettivamente per il modello (M.i.o) e (M.ii) le seguenti:

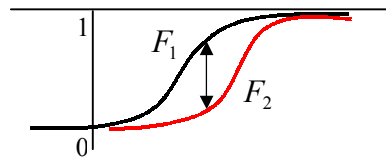
$$(M.i.o) \quad \Rightarrow \quad H_0 : \{X_1 \stackrel{d}{=} X_2\} \text{ contro } H_1 : \{X_1 \stackrel{d}{=} (X_2 - \delta)\}$$

$$(M.ii) \quad \Rightarrow \quad H_0 : \{X_1 \stackrel{d}{=} X_2\} \text{ contro } H_1 : \{X_1 \stackrel{d}{<} X_2\}$$

In accordo alla procedura multi-aspect descritta nel capitolo 6.4, procediamo con la prima fase, vale a dire la definizione degli aspetti e la scelta delle statistiche test corrispondenti.

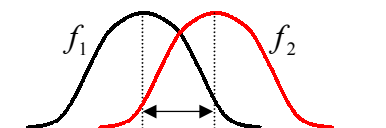
#### 8.3.1 PRIMA FASE: DEFINIZIONE DEGLI ASPETTI

*distance-aspect:* questo aspetto si basa su una misura della distanza fra le due funzioni di ripartizione empiriche, e quindi confrontano direttamente le due distribuzioni, senza ricorrere ad una loro sintesi;



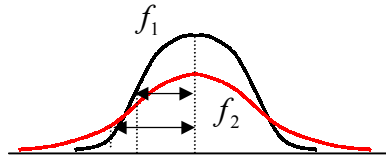
**Figura 8.3a:** confronto fra due funzioni di ripartizione

*location-aspect:* questo aspetto, sintetizzando le distribuzioni sottostanti, si basa sul confronto fra due indici di posizione;



**Figura 8.3b:** confronto fra due indici di posizione (per semplicità illustrativa, le due distribuzioni sono omoschedastiche)

*scatter-aspect*: questo aspetto, sintetizzando le distribuzioni sottostanti, si basa sul confronto fra due misure di dispersione;



**Figura 8.3c:** confronto fra due misure di dispersione (per semplicità illustrativa, le due distribuzioni hanno medie uguali)

In corrispondenza ai due modelli di risposta presentati nei paragrafi 8.2.1 e 8.2.2, si possono rispettivamente considerare i seguenti aspetti:

$$(M.i.o) \Rightarrow H_0 : \{X_1 \stackrel{d}{=} X_2\} \text{ contro } H_1 : \{X_1 \stackrel{d}{=} (X_2 - \delta)\} \quad \left\{ \begin{array}{l} \textit{distance - aspect} \\ \textit{location - aspect} \end{array} \right.$$

$$(M.ii) \Rightarrow H_0 : \{X_1 \stackrel{d}{=} X_2\} \text{ contro } H_1 : \{X_1 < X_2\} \quad \left\{ \begin{array}{l} \textit{distance - aspect} \\ \textit{location - aspect} \\ \textit{scatter - aspect} \end{array} \right.$$

### 8.3.2 PRIMA FASE: SCELTA DELLE STATISTICHE TEST PARZIALI DI PERMUTAZIONE

Seguendo i criteri del paragrafo 6.6, capitolo 6, si considereranno le statistiche test presenti in letteratura appropriate per ciascun aspetto considerato, e se ne presenterà la controparte di permutazione dimostrandone la non distorsione e consistenza, in modo da effettuare una *scelta ragionevole*. Nelle tabelle sottostanti viene illustrata la prima fase della procedura multi-aspect, rispettivamente per i modelli (M.i.o) e (M.ii):

<b>Tabella 8.1-(M.i.o):</b> Prima fase		
<i>aspetti considerati</i>	<i>statistiche test</i>	<i>sotto-ipotesi</i>
<i>distance-aspect</i>	$T_{KS}^*$	$\Rightarrow H_{0KS} : \{F_1 = F_2\} \text{ contro } H_{1KS} : \{F_1 > F_2\}$
	$T_{AD}^*$	$\Rightarrow H_{0AD} : \{F_1 = F_2\} \text{ contro } H_{1AD} : \{F_1 > F_2\}$
<i>location-aspect</i>	$T_a^*$	$\Rightarrow H_{0a} : \{E(X_1) = E(X_2)\} \text{ contro } H_{1a} : \{E(X_1) < E(X_2)\}$
	$T_{Me}^*$	$\Rightarrow H_{0Me} : \{Me(X_1) = Me(X_2)\} \text{ contro } H_{1Me} : \{Me(X_1) < Me(X_2)\}$

**Tabella 8.1-(M.ii):** Prima fase

<i>aspetti considerati</i>	<i>statistiche test</i>	<i>sotto-ipotesi</i>
<i>distance-aspect</i>	$T_{KS}^*$	$\Rightarrow H_{0KS} : \{F_1 = F_2\}$ contro $H_{1KS} : \{F_1 > F_2\}$
	$T_{AD}^*$	$\Rightarrow H_{0AD} : \{F_1 = F_2\}$ contro $H_{1AD} : \{F_1 > F_2\}$
<i>location-aspect</i>	$T_a^*$	$\Rightarrow H_{0a} : \{E(X_1) = E(X_2)\}$ contro $H_{1a} : \{E(X_1) < E(X_2)\}$
	$T_{Me}^*$	$\Rightarrow H_{0Me} : \{Me(X_1) = Me(X_2)\}$ contro $H_{1Me} : \{Me(X_1) < Me(X_2)\}$
<i>scatter-aspect</i>	$T_s^*$	$\Rightarrow H_{0s} : \{E(X_1^2) = E(X_2^2)\}$ contro $H_{1s} : \{E(X_1^2) < E(X_2^2)\}$

#### 8.4 STATISTICHE TEST UTILIZZATE PER IL *DISTANCE-ASPECT*

Il confronto di due distribuzioni continue rappresenta un tipico problema di *goodness of fit*, il quale dà origine a una famiglia di soluzioni le cui maggiormente note e rappresentative sono le statistiche test di Kolmogorov-Smirnov e di Cramèr-von Mises. La prima rientra nella classe che viene definita del “*supremum*” mentre la seconda nella classe definita “*quadratic*” (D’Agostino e Stephens, 1986) Entrambe sono funzione della distanza fra la funzione di ripartizione osservata e la corrispondente funzione di ripartizione ipotizzata sotto  $H_0$ ; possono inoltre essere estese al caso di due o più campioni. In mancanza di informazioni *a priori* relative alle distribuzioni sottostanti, si considereranno due statistiche test, quella di Kolmogorov-Smirnov e quella di Anderson-Darling, poiché la prima non è sensibile alle code delle distribuzioni, mentre la seconda è “pesata” risultando una statistica su componenti standardizzate.

*Osservazione 8.2:* I test tipo *goodness of fit* sono consistenti contro *qualsiasi* differenza tra le due funzioni di ripartizione; acquisendo questa protezione contro tutte le differenze, sacrificano potenza rispetto a sottoclassi di alternative (come spostamenti di posizione o differenze in dispersione)(Hollander e Wolfe, 1999).

##### 8.4.1 TEST PARZIALE DI PERMUTAZIONE $T_{KS}$

$$T_{KS}^* = \sup_{i \in \{1, \dots, n_1\}} \{F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)\}$$

$$H_{0KS} : \{F_1(x) = F_2(x), \forall x \in \mathfrak{R}^1\} \text{ contro } H_{1KS} : \{F_1(x) > F_2(x), \text{ per almeno un valore di } x \in \mathfrak{R}^1\}$$

dove  $F_j(x) = \#(X_{ji} \leq x) / n_j$ ,  $j = 1, 2$ , rappresenta la funzione di ripartizione empirica del  $j$ -esimo campione;  $\hat{F}(x) = \{[n_1 \cdot F_1(x) + n_2 \cdot F_2(x)] / n\}$  rappresenta la funzione di ripartizione empirica del campione congiunto e  $X_{1(i)}^*$  è l' $i$ -esima statistica ordinata del primo campione permutato  $X_{1(1)}^* \leq \dots \leq X_{1(n_1)}^*$ . In questa espressione, si è considerata l'alternativa unilaterale corrispondente alla sotto-ipotesi alternativa  $H_{1KS} : \{F_1(x) > F_2(x), \text{ per almeno un valore di } x \in \mathfrak{R}^1\}$ .

*Non distorsione:* Si consideri  $T_{KS}^*$  come una trasformazione monotona e non decrescente di  $X$ , e quindi  $\varphi(X + \delta_2) \geq \varphi(X + \delta_1)$  con  $\delta_1 = 0$  e  $\delta_2 > 0$ .

*Controparte in letteratura* Il test a due campioni di Kolmogorov-Smirnov, richiede che i dati consistano in due campioni casuali indipendenti di numerosità  $n_1$  e  $n_2$  e che le variabili casuali  $X_1$  e  $X_2$  siano assunte continue, al fine di garantire l'esattezza del test. Se le variabili casuali sono discrete, il test rimane valido, seppur diventi conservativo. Siano  $F_j(x) = \#(X_{ji} \leq x) / n_j$ ,  $j = 1, 2$ , le due funzioni di distribuzione empirica relative al  $j$ -esimo campione; la statistica test viene definita nella versione unilaterale e bilaterale come segue:

$$H_1 : \{F_1(x) \neq F_2(x), \text{ per almeno un valore di } x \in \mathfrak{R}^1\} \Rightarrow KS = \sup_x |F_1(x) - F_2(x)|$$

$$H_1 : \{F_1(x) > F_2(x), \text{ per almeno un valore di } x \in \mathfrak{R}^1\} \Rightarrow KS^+ = \sup_x \{F_1(x) - F_2(x)\}$$

La statistica di Kolmogorov-Smirnov misura la discrepanza fra le due funzioni di ripartizione empiriche, ma spesso non è abbastanza sensibile alle deviazioni nelle code delle distribuzioni (Conover, 1999).

#### 8.4.2 TEST PARZIALE DI PERMUTAZIONE $T_{AD}$

$$T_{AD}^* = \sum_{i=1}^{n_1} \frac{[F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)]}{\sqrt{[\hat{F}(X_{1(i)}^*)] \cdot [1 - (\hat{F}(X_{1(i)}^*))]}}$$

$$H_{0AD} : \{F_1(x) = F_2(x), \forall x \in \mathfrak{R}^1\} \text{ contro } H_{1AD} : \{F_1(x) > F_2(x), \text{ per almeno un valore di } x \in \mathfrak{R}^1\}$$

dove le notazioni sono le stesse del paragrafo 8.3.1.



*Non distorsione:* Si consideri  $T_{AD}^*$  come una trasformazione monotona e non decrescente di  $X$ , e quindi  $\varphi(X + \delta_2) \geq \varphi(X + \delta_1)$  con  $\delta_1 = 0$  e  $\delta_2 > 0$ .

*Controparte in letteratura:* Il test a due campioni di Cramér-von Mises, richiede che i dati consistano in due campioni casuali indipendenti di numerosità  $n_1$  e  $n_2$  e che le variabili casuali  $X_1$  e  $X_2$  siano assunte continue, al fine di garantire l'esattezza del test. Se le variabili casuali sono discrete, il test rimane valido, seppur diventi conservativo.

Siano  $F_j(x) = \#(X_{ji} \leq x) / n_j$ ,  $j = 1, 2$ , le due funzioni di distribuzione empirica relative al  $j$ -esimo campione; la statistica test viene definita nella versione bilaterale come segue:

$$H_1 : \{F_1(x) \neq F_2(x), \text{ per almeno un valore di } x \in \mathfrak{R}^1\} \Rightarrow$$

$$CM = \frac{(n_1 \cdot n_2)}{(n_1 + n_2)} \cdot \left\{ \sum_{i=1}^{n_1} [F_1(X_{1i}) - F_2(X_{1i})]^2 + \sum_{i=1}^{n_2} [F_1(X_{2i}) - F_2(X_{2i})]^2 \right\}$$

Si noti che la statistica di Cramér-von Mises appartiene alla *Cramér-von Mises family*, definita come:

$$CMF = \int_{-\infty}^{+\infty} \sum_{j=1}^2 n_j [F_j(x) - \hat{F}(x)]^2 \cdot \psi(x) d\hat{F}(x)$$

con  $\psi(x)$  una funzione opportuna che fornisce i pesi alla differenza quadratica; pertanto per  $CM$  avremo  $\psi(x) = 1$ . Dal momento che  $\text{Var}[F_n(t)] = \{F(t) \cdot [1 - F(t)]\} / n$ , la statistica test con  $\psi(x) = \hat{F}(x) \cdot [1 - \hat{F}(x)]$  prende il nome di Anderson-Darling. In accordo con Miller (1966), si considererà la statistica  $T_{AD}^*$  nei casi di alternative con dominanza stocastica, poiché “la estrema generalità dell'ipotesi  $H_1$  posta in alternativa all'ipotesi  $H_0$  rende i test scarsamente potenti e questo è il motivo delle perplessità manifestate da taluni autori sull'utilizzo non sufficientemente meditato dei test medesimi”.

## 8.5 STATISTICHE TEST UTILIZZATE PER IL LOCATION-ASPECT

Nel problema del confronto fra due indici di posizione è di interesse sapere qual'è l'indicatore appropriato per rilevare gli effetti del trattamento.

In generale, quanto più un indice statistico riesce a velare gli inquinamenti della distribuzione, ossia quanto meno è influenzato da dati anomali<sup>4</sup>, tanto più è detto robusto.

---

<sup>4</sup> outliers;

Più un indice è robusto, tanto minore è la sua capacità informativa perché tanto minore risulta l'effetto dei dati anomali sul valore dell'indice. Supponiamo, per esempio, che il termine massimo di una distribuzione sia molto grande, sia cioè anomalo: la media aritmetica della distribuzione risulta maggiore di quanto sarebbe stata se il termine massimo non fosse uscito dalla normalità; la mediana invece non risente della variazione della valore del termine massimo.

Pertanto, in mancanza di informazioni *a priori* relative alle distribuzioni sottostanti, si considereranno due statistiche test, una basata sul confronto tra le medie campionarie e l'altra tra le mediane campionarie.

### 8.5.1 TEST PARZIALE DI PERMUTAZIONE $T_a$

$$T_a^* = \sum_{i=1}^{n_2} X_{2i}^*$$

$$H_{0a} : \{E(X_1) = E(X_2)\} \text{ contro } H_{1a} : \{E(X_1) < E(X_2)\}$$

*Non distorsione:* Si è dimostrato (Capitolo 3, paragrafo 3.6), che, sotto il modello (M.i.o), per ogni  $\alpha \in \Lambda_{X(0)}$ , per ogni distribuzione  $P$  e per tutti i possibili insiemi di dati  $\mathbf{X} \in \Omega^n$ , la statistica  $T_a^*$  dà origine a un test di permutazione condizionatamente non distorto, cioè  $\Pr\{\lambda(X(\delta)) \leq \alpha \mid \Omega_{X(\delta)}^n\} \geq \alpha$ . In Pesarin (2001), la proprietà di non distorsione condizionata è generalizzata a problemi multi-campioni e a qualunque forma di dominanza stocastica, in accordo al modello di risposta (M.ii).

*Consistenza:* Considerando il modello (M.ii), la struttura di permutazione del test  $T_a^*$  nell'alternativa assume la forma  $\{T_a^*(0) + \sum_{i=1}^{n_2} \Delta_{2i}^* - \sum_{i=1}^{n_2} \Delta_{2i}\}$  dove  $T_a^*(0)$  dipende essenzialmente dai termini casuali sotto  $H_0$ , mentre la quantità  $\sum_{i=1}^{n_2} \Delta_{2i}^* - \sum_{i=1}^{n_2} \Delta_{2i}$  dipende essenzialmente dagli effetti trattamento (la quantità non negativa  $\sum_{i=1}^{n_2} \Delta_{2i}$  rappresenta l'effetto globale di permutazione dovuto all'assegnazione delle unità al secondo gruppo attraverso una permutazione casuale \* tale che  $T_a^*(\Delta) = T_a^*(0) + \sum_{i=1}^{n_2} \Delta_{2i}^*$ . In generale, quando la struttura di permutazione di una statistica assume questa forma, allora il test è permutazionalmente esatto, non distorto e consistente, a condizione che la quantità  $\sum_{i=1}^{n_2} \Delta_{2i}^* - \sum_{i=1}^{n_2} \Delta_{2i}$  diverga almeno debolmente quando la numerosità campionaria tende all'infinito. Ciò si verifica quando  $X$  ammette finito il momento primo (Pesarin, 2001).

Si vuole sottolineare che utilizzando la statistica test parziale di permutazione  $T_a^*$  viene garantita la consistenza del test combinato che la considera come componente, in base all'assunzione (Ar.2) e al rilassamento del teorema 6.1, 6.3.1.

*Controparte in letteratura:* La statistica test  $T_a^*$  risulta permutazionalmente equivalente al confronto standardizzato di due medie campionarie; è noto che sotto un modello (M.i.o) con  $P = N(\mu, \sigma^2)$  il test  $t$  di Student è il test simile uniformemente più potente per verificare l'ipotesi nulla di posizioni uguali contro l'ipotesi alternativa unilaterale (Lehmann, 1986). Quindi sotto queste condizioni, o più generalmente quando le distribuzioni sottostanti sono simmetriche e in assenza di dati anomali, il test  $T_a^*$  rappresenta una scelta ragionevole.

### 8.5.2 TEST PARZIALE DI PERMUTAZIONE $T_{Me}$

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^*$$

$$H_{0Me} : \{Me(X_1) = Me(X_2)\} \text{ contro } H_{1Me} : \{Me(X_1) < Me(X_2)\}$$

$$\text{dove } \tilde{M}_j^* \text{ è definito come } \tilde{M}_j^* = \begin{cases} X_{j([n_j+1]/2)}^* & \text{se } n_j \text{ è dispari} \\ 0.5 \cdot \{X_{j(n_j/2)}^* + X_{j([n_j/2]+1)}^*\} & \text{se } n_j \text{ è pari} \end{cases} \quad j = 1, 2$$

dove  $\tilde{M}_j^*$  ( $j = 1, 2$ ) rappresenta la mediana calcolata sull' $j$ -esimo campione,  $X_{j(i)}^*$  è la  $i$ -esima

( $i = 1, \dots, n_j$ ) statistica ordinata del  $j$ -esimo campione permutato:  $X_{j(1)}^* \leq \dots \leq X_{j(n_j)}^*$ .

*Non distorsione:* Si consideri il modello (M.i.o); per semplicità di esposizione siano le due numerosità campionarie  $n_1$  e  $n_2$  dispari (si può dimostrare che i risultati valgono per  $n_j$  pari o dispari,  $j = 1, 2$ ); i valori osservati di  $T_{Me}$  sotto  $H_0$  e  $H_1$  risultano rispettivamente:

$$T_{Meo}(0) = X_{2([n_2+1]/2)} - X_{1([n_1+1]/2)}$$

$$T_{Meo}(\delta) = X_{2([n_2+1]/2)} + \delta - X_{1([n_1+1]/2)} = T_{Meo}(0) + \delta$$

Si consideri una generica permutazione:

$$T_{Meo}^*(0) = X_{2([n_2+1]/2)}^* - X_{1([n_1+1]/2)}^*$$

$$\begin{aligned} T_{Meo}^*(\delta) &= X_{2([n_2+1]/2)}^* + \delta \cdot [\mathbf{I}(X_{2([n_2+1]/2)}^* \in X_2)] - X_{1([n_1+1]/2)}^* - \delta \cdot [\mathbf{I}(X_{1([n_1+1]/2)}^* \in X_2)] \\ &= T_{Me}^*(0) + \delta \cdot [\mathbf{I}(X_{2([n_2+1]/2)}^* \in X_2) - \mathbf{I}(X_{1([n_1+1]/2)}^* \in X_2)] \end{aligned}$$

Quindi si ottiene:

$$\begin{aligned} \lambda(\mathbf{X}(\delta)) &= \Pr\{T_{Me}^*(\delta) \geq T_{Meo}^*(\delta) \mid \Omega_{\mathbf{X}(\delta)}^n\} \\ &= \Pr\{T_{Me}^*(0) + \delta \cdot [\mathbf{I}(X_{2([n_2+1]/2)}^* \in X_2) - \mathbf{I}(X_{1([n_1+1]/2)}^* \in X_2)] - \delta \geq T_{Meo}^*(0) \mid \Omega_{\mathbf{X}(0)}^n\} \\ &\leq \Pr\{T_{Me}^*(0) \geq T_{Meo}^*(0) \mid \Omega_{\mathbf{X}(0)}^n\} = \lambda(\mathbf{X}(0)) \end{aligned}$$

poiché  $\delta \cdot [\mathbf{I}(X_{2([n_2+1]/2)}^* \in X_2) - \mathbf{I}(X_{1([n_1+1]/2)}^* \in X_2) - 1] \leq 0$ .

*Controparte in letteratura:* Con riferimento al test  $T_b$  discusso nel capitolo 6, esempio 4, viene presentato in letteratura il test delle mediane, progettato per esaminare se più campioni provengono da popolazioni aventi la stessa mediana. In realtà, il test della mediana è soltanto una speciale applicazione del test Chi-quadrato con i margini totali fissati. Per verificare se più popolazioni hanno la stessa mediana, un campione casuale viene estratto da ciascuna popolazione. Si costruisce una tabella di contingenza  $2 \times c$  ( $c$  è il numero dei campioni), e le due entità dell' $i$ -esima colonna ( $i = 1, \dots, c$ ) sono il numero delle osservazioni dell' $i$ -esimo campione che risultano superiori e inferiori alla “*grand median*” (la mediana di tutte le osservazioni riunite). L'usuale test Chi-quadrato viene quindi applicato alla tabella di contingenza (Conover, 1999).

Il test  $T_{Me}$  può essere semplicemente considerato come il confronto fra due statistiche ordinate, le mediane campionarie, e pertanto ci si attende che si comporti meglio del test basato sul confronto fra le medie campionarie in presenza di dati asimmetrici e/o di dati anomali o provenienti da distribuzioni con valor medio inesistente.

## 8.6 TEST PARZIALE DI PERMUTAZIONE PER LO SCATTER-ASPECT

$$T_s^* = \sum_{i=1}^{n_2} (X_{2i}^*)^2$$

$$H_{0s} : \{E(X_1^2) = E(X_2^2)\} \text{ contro } H_{1s} : \{E(X_1^2) < E(X_2^2)\}$$

*Non distorsione:* Per il modello (M.ii), si consideri una generica permutazione nella quale  $v^*$  elementi sono scambiati casualmente fra i due gruppi, cosicché, si ottiene:

$$T_s^*(\Delta) = \sum_{i=1}^{n_2} (\mu + \Delta_{2i}^* + Z_{2i}^*)^2 \geq T_s^*(0) = (\mu + Z_{2i}^*)^2$$

poiché in  $T_s^*(\Delta)$  vi sono  $v^*$  elementi con  $\Delta_{2i}^* = 0$  e  $(n_2 - v^*)$  elementi con  $\Delta_{2i}^* \geq 0$ .

*Controparte in letteratura:* Il test può contenere informazioni sugli effetti di secondo ordine, come illustrato in Pesarin (2001).

### 8.7 SECONDA E TERZA FASE: STRUTTURAZIONE DEGLI ASPETTI E SCELTA DELLE FUNZIONI DI COMBINAZIONE

Si procede ora alla strutturazione degli aspetti rispettivamente per i modelli (M.i.o) e (M.ii) (tabelle 8.2-(M.i.o) e 8.2-M(ii)):

**Tabella 8.2-(M.i.o):** Seconda fase

<i>aspetto</i>	<i>sotto-ipotesi associate</i>	<i>struttura</i>	<i>ipotesi nulla associata</i>
<i>distance-aspect</i>	$H_{0d} : \{F_1 = F_2\}$		$H_0 : \left\{ \begin{array}{c} [F_1 = F_2] \\ \cap \\ \left[ \begin{array}{c} (E(X_1) = E(X_2)) \cap \\ (Me(X_1) = Me(X_2)) \end{array} \right] \end{array} \right\}$
<i>location-aspect</i>	$H_{0l} : \left\{ \begin{array}{c} (E(X_1) = E(X_2)) \\ \cap \\ (Me(X_1) = Me(X_2)) \end{array} \right\}$		

Per il modello (M.i.o) si è costretti a strutturare “direttamente” poiché vengono considerati solo due aspetti, a cui verrà assegnata la stessa importanza;

**Tabella 8.2-(M.ii):** Seconda fase

<i>aspetto</i>	<i>sotto-ipotesi associate</i>	<i>struttura</i>	<i>ipotesi nulla associata</i>
<i>distance-aspect</i>	$H_{0d} : \{F_1 = F_2\}$		$H_0 : \left\{ \begin{array}{c} [F_1 = F_2] \\ \cap \\ \left[ \begin{array}{c} (E(X_1) = E(X_2)) \cap \\ (Me(X_1) = Me(X_2)) \end{array} \right] \cap \\ (E(X_1^2) = E(X_2^2)) \end{array} \right\}$
<i>location-aspect</i>	$H_{0l} : \left\{ \begin{array}{c} (E(X_1) = E(X_2)) \\ \cap \\ (Me(X_1) = Me(X_2)) \end{array} \right\}$		
<i>scatter-aspect</i>	$H_{0s} : \{E(X_1^2) = E(X_2^2)\}$		

per il modello (M.ii), invece, si struttura in una fase intermedia il *location-aspect* con lo *scatter-aspect*, dando origine a un macro aspetto che si può denominare *jls*<sup>5</sup>-*aspect*, che successivamente verrà combinato con il *distance-aspect*. Di conseguenza, si assegnerà lo stesso “peso” all’aspetto che considera l’analisi congiunta della posizione e della dispersione e a quello che valuta la distanza tra le due funzioni di ripartizione empiriche.

La terza fase viene effettuata in due passi, e comporta le scelte delle funzioni di combinazione rispettivamente all’interno degli aspetti (tabelle 8.3-i) e fra gli aspetti (tabelle 8.3-ii);

<b>Tabella 8.3-(M.i.o)-i:</b> terza fase, primo passo					
<i>aspetto</i>	<i>statistiche test</i>	<i>p-values dei test</i>	<i>sotto-ipotesi dell'aspetto</i>	<i>test combinato associato all'aspetto</i>	<i>p-value aspetto</i>
<i>distance-aspect</i>	$(T_{KS}, T_{AD})$	$(\lambda_{KS}, \lambda_{AD})$	$H_{0d} : \begin{pmatrix} H_{KS} \\ \cap \\ H_{AD} \end{pmatrix}$	$T_d'' = \psi_T(\lambda_{KS}, \lambda_{AD})$	$\lambda_d''$
<i>location-aspect</i>	$(T_a, T_{Me})$	$(\lambda_a, \lambda_{Me})$	$H_{0l} : \begin{pmatrix} H_a \\ \cap \\ H_{Me} \end{pmatrix}$	$T_l'' = \psi_T(\lambda_a, \lambda_{Me})$	$\lambda_l''$

<b>Tabella 8.3-(M.ii)-i:</b> terza fase, primo passo					
<i>aspetto</i>	<i>statistiche test</i>	<i>p-values dei test</i>	<i>sotto-ipotesi dell'aspetto</i>	<i>test combinato associato all'aspetto</i>	<i>p-value aspetto</i>
<i>distance-aspect</i>	$(T_{KS}, T_{AD})$	$(\lambda_{KS}, \lambda_{AD})$	$H_{0d} : \begin{pmatrix} H_{KS} \\ \cap \\ H_{AD} \end{pmatrix}$	$T_d'' = \psi_T(\lambda_{KS}, \lambda_{AD})$	$\lambda_d''$
<i>location-aspect</i>	$(T_a, T_{Me})$	$(\lambda_a, \lambda_{Me})$	$H_{0l} : \begin{pmatrix} H_a \\ \cap \\ H_{Me} \end{pmatrix}$	$T_l'' = \psi_T(\lambda_a, \lambda_{Me})$	$\lambda_l''$
<i>scatter-aspect</i>	$(T_s)$		$H_{0s}$		$\lambda_s$

Per combinare le statistiche test parziali di permutazione del *distance-aspect* ( $T_{KS}$  e  $T_{AD}$ ) e del *location-aspect* ( $T_a$  e  $T_{Me}$ ) si è scelta la funzione di combinazione di Tippett poiché si vuole preferire per ciascun aspetto una delle due statistiche, che possono essere considerate “integrative”. Infatti, la funzione di combinazione di Tippett opererà per  $T_{AD}$  o

<sup>5</sup> *j.l.s.* : joint location (and) scatter;

per  $T_{KS}$  rispettivamente se si vuole dare più o meno attenzione alle code della distribuzione e se si vuole fare un uso più o meno effettivo dei dati<sup>6</sup>, mentre opererà per  $T_a$  o per  $T_{Me}$  rispettivamente per distribuzioni simmetriche oppure per distribuzioni asimmetriche e/o in presenza di dati anomali.

**Tabella 8.3-(M.i.o)-ii:** terza fase, secondo passo

aspetto	struttura	combinazione globale	p-value globale
distance- aspect location- aspect	 $\{distance \cap location\}$	$T_{MA}''' = \psi_T(\lambda_d'', \lambda_l'')$	$\lambda_{MA}'''$

Nel modello (M.i.o) per combinare gli aspetti *distance* e *location* si è scelta la funzione di combinazione di Tippett, poiché si vuole dar la preferenza all'aspetto che meglio interpreti il modello.

**Tabella 8.3-(M.ii)-ii:** terza fase, secondo passo

aspetto	struttura	combinazione intermedia	p-value	combinazione globale	p-value globale
distance- aspect location- aspect scatter- aspect	 $\{[distance] \cap [location \cap scatter]\}$	$T_{jls}''' = \psi_F(\lambda_l'', \lambda_s'')$	$\lambda_{jls}'''$	 $T_{MA}'''' = \psi_T(\lambda_d'', \lambda_{jls}''')$	$\lambda_{MA}''''$

Nel modello (M.ii), la combinazione intermedia degli aspetti *location* e *scatter* è ottenuta attraverso la funzione di combinazione di Fisher, in quanto si presume che entrambi siano similmente informativi: si noti che  $T_a$  e  $T_s$  potrebbero essere interpretate come se fossero lo stesso test parziale applicato alla variabile  $X$  e ad una sua trasformazione  $Y = \varphi(X) = X^2$ ; sebbene quest'ultima componente possa sembrare non informativa, essendo associata a  $X$  con probabilità uno, è utile notare che il relativo test parziale può contenere informazioni sugli effetti di secondo ordine.

<sup>6</sup> *it seems* (Cramèr-von Mises) *to make more effective use of the data*, Conover (1999);

Si noti che attraverso la combinazione intermedia nell'aspetto *jls* si assegnano “pesi” dimezzati ai due aspetti *location* e *scatter*, in modo da non sbilanciare eccessivamente l'inferenza globale sulle loro statistiche test.

Ottenuto in questo modo il macro-aspetto *jls*, esso sarà combinato con il *distance-aspect* attraverso la funzione di combinazione di Tippett: si preferirà quindi o l'aspetto che utilizza test parziali che sono funzioni di statistiche ordinate oppure quello che, seppur contenendo la statistica test relativa alla differenza fra le mediane campionarie, dà maggior importanza ai test parziali che sono funzioni delle osservazioni.

## 8.8 STUDIO COMPARATIVO VIA SIMULAZIONE PER IL MODELLO CON EFFETTI ADDITTIVI FISSI

Prendendo in considerazione lo studio di simulazione presentato da Marozzi(2003) e brevemente discusso nell'Esempio 6.4, Capitolo 6, si possono riassumere nella seguente tabella 8.4 le considerazioni che hanno motivato il test bi-aspect proposto:

**Tabella 8.4:** Criteri seguiti nella costruzione del test bi-aspect

Box e Anderson (1955) suggeriscono un principio per stabilire la bontà di un test: <i>per adempiere alle necessità dello sperimentatore, i criteri statistici devono</i>	
↓	↓
(1) <i>essere sensibili ai cambiamenti dei fattori specifici verificati</i>	(2) <i>essere insensibili a cambiamenti dovuti a fattori estranei di una grandezza plausibile a verificarsi nella pratica</i>
↓	↓
potenza e non distorsione	robustezza
un test parametrico è quasi sempre sviluppato al fine di soddisfare (1), a condizione che le proprie assunzioni vengano raggiunte nella pratica	un test non parametrico è innatamente robusto, dal momento che richiede assunzioni meno specifiche, che comunque devono essere soddisfatte
si propone un test con buone prestazioni generali, che possa quindi essere utilizzato senza la preoccupazione di quale sia la distribuzione sottostante ai dati, la quale di solito risulta non nota.	
↓	
test bi-aspect	
aspetto “numerico” $T_a = \sum_{i=1}^{n_2} X_{2i}$	aspetto “categoriale” $T_b = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i} > \tilde{M})$
↓	↓
per dati simmetrici per i quali la media è una statistica sensibile nel rilevare spostamenti di posizione	per dati asimmetrici e in presenza di <i>outliers</i> per i quali la mediana è una statistica sensibile nel rilevare spostamenti di posizione
↓	
combinazione non parametrica di test dipendenti	
↓	
il test globale $T_{ab}$ ottenuto attraverso la funzione di combinazione di Tippett avrà prestazioni prossime all'aspetto (test parziale di permutazione) maggiormente adeguato alla distribuzione di popolazione sottostante non nota;	



*Osservazione 8.3:* Con riferimento al criterio (2), vale la pena notare che:

- L'approccio parametrico si basa su assunzioni specifiche riguardanti la natura della distribuzione sottostante: spesso si assume che la sua forma sia nota ad eccezione di pochi parametri. Nell'approccio non parametrico, invece, le assunzioni specifiche sulla distribuzione vengono rimpiazzate da una molto generale: può essere semplicemente richiesto che la popolazione sia continua. Si osservi che condizionare ad un'insieme di statistiche sufficienti per  $P$  rende il test di permutazione *distribution-free* e non parametrico;
- Un test parametrico è esatto solo se la distribuzione sottostante è quella sulla quale il test è basato. In questo caso, il test è spesso il miglior test disponibile. Ma se la distribuzione sottostante è differente, il test parametrico perde queste proprietà e la sua prestazione può essere scarsa. Al contrario, una condizione sufficiente per una statistica test  $T$  per essere permutazionalmente esatta è che la sua distribuzione sotto  $H_0$  possa essere espressa solo in termini di permutazioni degli errori casuali  $\mathbf{Z}$ , a condizione che essi siano scambiabili (Pesarin, 2001, Remark 2, 3.3.2). Quindi, considerando un modello ad effetti additivi fissi (M.i.o), questa condizione viene soddisfatta ad esempio dalla statistica test  $T_a = \sum_{i=1}^{n_2} X_{2i}$ : si consideri  $v^*$  dati sono stati scambiati casualmente fra i due gruppi, quindi la struttura di permutazione risulta  $T_a^* = \sum_{i=1}^{n_2} X_{2i}^* = n_2\mu + (n_2 - v^*)\delta_2 + \sum_{i=1}^{n_2} \sigma \cdot Z_{2i}^*$  pertanto, se e solo se  $H_0$  risulta vera,  $T_a^*$  dipende solo dagli errori scambiabili, poiché le altre quantità si semplificano o sono costanti. Il test combinato risulterà esatto se se lo sono tutti i test parziali di permutazione.

Per valutare le prestazioni dei test che verranno utilizzati, sono state condotte simulazioni Monte Carlo al fine di stimarne la potenza con il numero di proporzioni di rifiuto dell'ipotesi nulla quando essa risulta falsa, e l'errore di I tipo con il numero di proporzioni di rifiuto dell'ipotesi nulla quando essa risulta vera. Verranno effettuate  $MC = 2000$  simulazioni Monte Carlo,  $B = 1000$  permutazioni (ricampionamenti condizionati CMC), con un livello di significatività  $\alpha$  fissato al 5%. La definizione di livello di significatività osservato è quella descritta nel capitolo 6, osservazione 6.7.

### 8.8.1 MODELLO E IPOTESI

Siano  $X_{11}, \dots, X_{1n_1}$  e  $X_{21}, \dots, X_{2n_2}$  due campioni casuali indipendenti ottenuti da popolazioni distribuite secondo il seguente modello con effetti additivi fissi (M.i.o):

$$\begin{aligned} X_{1i} &= Z_{1i} \\ X_{2i} &= \delta_2 + Z_{2i} \quad i = 1, \dots, n_j, j = 1, 2 \end{aligned}$$

dove  $\delta = \delta_2$  è l'effetto trattamento e  $Z_{ji}$  sono errori casuali scambiabili con distribuzione comune  $P$ . Si assume perciò che sotto  $H_1$  la distribuzione di  $X_2$  è spostata di una quantità  $\delta > 0$  rispetto a quella di  $X_1$ , così che  $H_1 : \{X_1 \stackrel{d}{=} (X_2 - \delta)\}$ . Con riferimento alle alternative unidirezionali, le ipotesi possono essere scritte in maniera equivalente come  $H_0 : \{\delta = 0\}$  e  $H_1 : \{\delta > 0\}$ .

Sono state considerate tre configurazioni con uguale numerosità campionaria:

$$(n_1, n_2) = (10, 10), (15, 15), (20, 20)$$

e quattro coppie con numerosità sbilanciata:

$$(n_1, n_2) = (15, 10), (20, 10), (10, 15), (10, 20).$$

Per stimare il tasso di errore di I tipo abbiamo considerato  $\delta = 0$ , e quindi sotto  $H_0$ .

Per stimare la potenza, abbiamo considerato valori positivi di  $\delta (> 0)$ , e quindi sotto  $H_1$ ) per i quali corrisponda approssimativamente una potenza di .25 per il test somma dei ranghi di Wilcoxon-Mann-Whitney con numerosità campionaria  $(n_1, n_2) = (10, 10)$ ,  $\alpha = .05$ , e con  $MC = 10\,000$ ; cioè

$\delta_2 = d : \Pr\{1 - \hat{\beta}(WMW | MC) | \delta_2 = d; H_1\} \cong 0.25$ . Il test di Wilcoxon-Mann-Whitney verrà quindi utilizzato come *standard* di comparazione.

### 8.8.2 STATISTICHE TEST UTILIZZATE

Verranno considerate le statistiche test di permutazione descritte nei paragrafi 8.3, 8.4 (test parziali) e 8.6 (test globali), relative al modello (M.i.o); per esigenze comparative, si considererà il test bi-aspect descritto nell'esempio 4, capitolo 6.

Inoltre, per effettuare ulteriori confronti, si considererà il test parametrico  $t$  di Student (unidirezionale) e i test non parametrici di Wilcoxon-Mann-Whitney ( $WMW$ ) e di Kolmogorov-Smirnov ( $KS^+$ , unidirezionale).

*Osservazione 8.4:* Per sormontare la sensibilità dei tassi di errore di I Tipo della statistica  $t$  di Student alle varie deviazioni dalla normalità, vengono utilizzate frequentemente procedure non parametriche basate sui ranghi. Queste ultime, tuttavia,

possono comportare la perdita di informazioni come conseguenza della sostituzione dei valori reali delle osservazioni con i loro ranghi (Keller-McNulty e Higgins, 1987).

### 8.8.3 DISTRIBUZIONI DI POPOLAZIONE CONSIDERATE

Verranno considerate le seguenti distribuzioni (tabella 8.5):

**Tabella 8.5:** distribuzioni considerate per il modello (M.i.o)

<i>distribuzioni simmetriche</i>		$E[P]$	$Me[P]$	$Var[P]$	<i>tabelle</i>
Normale standard	$N(0, 1)$	0	0	1	8.6 – 8.13
Laplace	$L(0, 1/\sqrt{2})$	0	0	1	8.6 – 8.13
Cauchy	$Cauchy(0, 1)$	-	0	-	8.6 – 8.13
Uniforme	$U(0, 2\cdot\sqrt{3})$	$\sqrt{3}$	$\sqrt{3}$	1	8.6, 8.7
Bimodale	$0.5\cdot N(-1.5, 1) + 0.5\cdot N(1.5, 1)$	0	0	0.5	8.6, 8.7
<i>t</i> di Student	<i>t</i> di Student(2 g.d.l.)	0	0	-	8.6, 8.7
<i>distribuzioni asimmetriche</i>		$E[P]$	$Me[P]$	$Var[P]$	<i>tabelle</i>
Half-Cauchy	Half-Cauchy(0,1)	-	-	-	8.6 – 8.13
Esponenziale	Exp(1)	1	0.7	1	8.6 – 8.13
contaminate <sup>7</sup> 10%	$0.9\cdot N(0,1) + 0.1\cdot N(1,100)$	0.1	0.1	1.81	8.6 – 8.13
Chi-quadrato n.c.	$\chi^2(0.25, 8)$	0.5	7.6	0.06	8.6, 8.7
contaminate <sup>7</sup> 10%	$0.9\cdot N(0,1) + 0.1\cdot N(9,1)$	0.9	0.9	0.82	8.6, 8.7
contaminate <sup>7</sup> 30%	$0.7\cdot N(0,1) + 0.3\cdot N(1,100)$	0.3	0.3	9.49	8.6, 8.7
contaminate <sup>7</sup> 30%	$0.7\cdot N(0,1) + 0.3\cdot N(9,1)$	2.7	2.7	0.82	8.6, 8.7

Il riferimento “tabelle” indica in quali tabelle di risultati viene considerata la distribuzione. Le distribuzioni esaminate sono le stesse considerate da Marozzi(2002) ad eccezione della distribuzione *t* di Student con due gradi di libertà e delle due distribuzioni contaminate con una percentuale di dati anomali provenienti da una Normale con media 9 e varianza 1<sup>(7)</sup>.

*Osservazione 8.5:* I due campioni casuali di numerosità variabile sono stati generati indipendentemente attraverso le funzioni *normrnd*, *trnd*, *unifrnd*, *exprnd*, *ncx2rnd* del *software* Matlab 6 rispettivamente per le distribuzioni normale, *t* di Student (quindi anche Cauchy), uniforme, esponenziale e chi-quadrato non centrale. Inoltre

- La distribuzione esponenziale doppia o di Laplace è ottenuta come la distribuzione esponenziale e quindi assegnando casualmente segni positivi o negativi (attraverso una prova di bernoulli equiprobabile) a ogni elemento del campione;
- La distribuzione bimodale è ottenuta come mistura di due distribuzioni normali ( $N(-1.5, 1)$  e  $N(1.5, 1)$ ) selezionate attraverso una prova di bernoulli equiprobabile;

<sup>7</sup> una mistura del genere è presente nel lavoro di Keller-McNulty e Higgins (1987);

- La distribuzione Half-Cauchy è stata generata come una Cauchy e cambiando il segno per ogni termine negativo del campione;
- La distribuzioni contaminate da una proporzione (10% e 30%) di dati anomali sono ottenute come mistura di due distribuzioni normali ( $N(0, 1)$  e  $N(1, 100) / N(9, 1)$ ) selezionate attraverso una prova di bernoulli appropriata alla percentuale di contaminazione;

Si noti che nel contesto dei metodi di permutazione non è necessario assumere che la media e la varianza delle distribuzioni sottostanti siano finite, come avviene infatti per le distribuzioni Cauchy e Half-Cauchy (non esistono momenti finiti) e  $t$  di Student con due gradi di libertà (esiste finito solo il momento di ordine uno). Questo può essere dovuto sia al principio del condizionamento sia al fatto che i test di permutazione sono sempre basati su un insieme finito di valori limitati, cosicché la media e la varianza dei dati osservati sono sempre finite.

**Tabella 8.6 :** Studio via simulazione, modello (M.i.o)

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)												
n <sub>1</sub> = n <sub>2</sub> = 10; α = 0.05; MC = 2000, B = 1000; δ <sub>2</sub> = 0.61, 0.5, 1.168, 0.61, 1.15, 0.86, 0.408, 2.3, 0.66, 0.74, 0.75, 1.5, 1.15;												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	T <sub>AD</sub>	T <sub>KS</sub>	T <sub>Me</sub>	T <sub>a</sub>	T <sub>b</sub>	T <sub>d</sub> <sup>''</sup>	T <sub>l</sub> <sup>''</sup>	T <sub>ab</sub> <sup>''</sup>	T <sub>MA</sub> <sup>''</sup>	t	KS <sup>+</sup>	WMW
Normale	.3715	.2940	.3425	.3785	.2500	.3585	.3660	.3730	.3625	.3835	.2110	.2555
Laplace	.3655	.3425	.3695	.3350	.2885	.3810	.3650	.3570	.3820	.3345	.2775	.2785
Cauchy	.3245	.3765	.3820	.2180	.3350	.3620	.3315	.2755	.3675	.1565	.3025	.2495
Uniforme	.3450	.2930	.3220	.3640	.2400	.3455	.3455	.3585	.3430	.3640	.2235	.2440
Bimodale	.3675	.2645	.2930	.3665	.1820	.3570	.3370	.3550	.3395	.3670	.1930	.2505
t di Student	.3505	.3420	.3735	.3050	.2925	.3635	.3705	.3310	.3800	.2840	.2665	.2565
Esponenziale	.4115	.3250	.2920	.2625	.2120	.3955	.3035	.2660	.3570	.2565	.2305	.2530
Chi-quadrato	.3845	.3045	.3300	.3525	.2435	.3725	.3545	.3545	.3645	.3535	.2210	.2620
Half-Cauchy	.3830	.3770	.2805	.1600	.2230	.3945	.2590	.1990	.3530	.1015	.2790	.2490
Contaminate' 10%	.3140	.3070	.3380	.2185	.2675	.3260	.3185	.2625	.3380	.0770	.2390	.2345
Contaminate'' 10%	.3680	.3195	.3520	.2360	.2890	.3670	.3370	.2775	.3520	.2030	.2420	.2545
Contaminate' 30%	.3510	.4860	.4580	.0915	.4375	.4415	.3665	.3035	.4420	.0625	.4110	.2930
Contaminate'' 30%	.3905	.3310	.2410	.1510	.2545	.3820	.2230	.1995	.3455	.1460	.2595	.2705

**Tabella 8.7** : studio via simulazione, modello (M.i.o)

H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)												
n <sub>1</sub> = n <sub>2</sub> = 10; α = 0.05; MC = 2000, B = 1000; δ <sub>2</sub> = 0;												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	T <sub>AD</sub>	T <sub>KS</sub>	T <sub>Me</sub>	T <sub>a</sub>	T <sub>b</sub>	T <sub>d</sub> <sup>''</sup>	T <sub>i</sub> <sup>''</sup>	T <sub>ab</sub> <sup>''</sup>	T <sub>MA</sub> <sup>'''</sup>	t	KS <sup>+</sup>	WMW
Normale	.0515	.0505	.0540	.0530	0.485	.0495	.0485	.0510	.0520	.0540	.0240	.0540
Laplace	.0605	.0545	.0560	.0540	.0570	.0545	.0540	.0550	.0520	.0540	.0290	.0530
Cauchy	.0475	.0450	.0490	.0515	.0445	.0465	.0490	.0490	.0505	.0260	.0275	.0510
Uniforme	.0420	.0410	.0460	.0425	.0455	.0455	.0450	.0415	.0435	.0440	.0250	.0455
Bimodale	.0490	.0500	.0505	.0545	.0450	.0455	.0545	.0545	.0520	.0530	.0265	.0575
t di Student	.0460	.0440	.0485	.0485	.0500	.0490	.0520	.0470	.0520	.0430	.0260	.0535
Esponenziale	.0455	.0420	.0415	.0430	.0415	.0435	.0440	.0455	.0405	.0400	.0220	.0455
Chi-quadrato	.0465	.0430	.0470	.0475	.0445	.0465	.0485	.0470	.0470	.0495	.0240	.0485
Half-Cauchy	.0580	.0555	.0535	.0515	.0495	.0535	.0545	.0525	.0545	.0255	.0270	.0535
Contaminate' 10%	.0415	.0430	.0380	.0390	.0440	.0405	.0400	.0360	.0420	.0170	.0235	.0420
Contaminate'' 10%	.0520	.0455	.0570	.0430	.0540	.0555	.0480	.0455	.0485	.0425	.0275	.0485
Contaminate' 30%	.0485	.0455	.0460	.0505	.0455	.0480	.0445	.0515	.0460	.0425	.0275	.0445
Contaminate'' 30%	.0490	.0470	.0465	.0470	.0460	.0490	.0485	.0460	.0440	.0495	.0235	.0440

$$T_{AD}^* = \sum_{i=1}^{n_1} \frac{[F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)]}{\sqrt{[\hat{F}(X_{1(i)}^*)] \cdot [1 - (\hat{F}(X_{1(i)}^*))]}}$$

$$T_{KS}^* = \sup_{i \in \{1, \dots, n_1\}} \{F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)\}$$

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^*$$

$$T_a^* = \sum_{i=1}^{n_2} X_{2i}^*$$

$$T_b^* = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i} > \tilde{M})$$

$$T_d^{''} = \psi_T(\lambda_{AD}, \lambda_{KS})$$

$$T_i^{''} = \psi_T(\lambda_{Me}, \lambda_a)$$

$$T_{ab}^{''} = \psi_T(\lambda_a, \lambda_b)$$

$$T_{MA}^{'''} = \psi_T(\lambda_d^{''}, \lambda_i^{''})$$

**Tabella 8.8** : studio via simulazione, modello (M.i.o)

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)												
$n_1 = 10, n_2 = 15; \alpha = 0.05; MC = 2000, B = 1000; \delta_2 = 0.61, 0.5, 1.168, 0.408, 0.74, 0.66;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.3850	.3340	.3430	.4025	.2275	.3660	.3840	.3840	.3720	.4030	.2925	.2620
Laplace	.4030	.4130	.4245	.3705	.3175	.4155	.4135	.4105	.4215	.3725	.3725	.2960
Cauchy	.3695	.4635	.4335	.2015	.3675	.4400	.3695	.4055	.4315	.1545	.4210	.2730
Esponenziale	.4885	.4235	.3670	.3035	.1685	.4730	.3710	.3050	.4225	.2895	.4010	.2930
Contaminate'	.3845	.3805	.4035	.2170	.2650	.3920	.3710	.3445	.3870	.0835	.3305	.2700
Half-Cauchy	.4895	.4750	.3840	.1815	.1625	.5120	.3490	.2265	.4460	.0910	.4680	.2975

H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)												
$n_1 = 10, n_2 = 15; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.0450	.0440	.0440	.0470	.0200	.0420	.0450	.0410	.0440	.0450	.0360	.0460
Laplace	.0470	.0470	.0460	.0470	.0260	.0440	.0490	.0460	.0440	.0470	.0390	.0450
Cauchy	.0590	.0550	.0500	.0550	.0290	.0570	.0540	.0560	.0610	.0320	.0470	.0590
Esponenziale	.0380	.0470	.0480	.0420	.0250	.0410	.0490	.0480	.0460	.0410	.0330	.0430
Contaminate'	.0410	.0500	.0440	.0350	.0260	.0430	.0400	.0400	.0390	.0170	.0330	.0270
Half-Cauchy	.0470	.0430	.0510	.0560	.0290	.0430	.0500	.0510	.0480	.0170	.0360	.0480

$$T_{AD}^* = \sum_{i=1}^{n_1} \frac{[F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)]}{\sqrt{[\hat{F}(X_{1(i)}^*)] \cdot [1 - (\hat{F}(X_{1(i)}^*))]}} \quad T_d'' = \psi_T(\lambda_{AD}, \lambda_{KS}) \quad T_{ab}'' = \psi_T(\lambda_a, \lambda_b)$$

$$T_{KS}^* = \sup_{i \in \{1, \dots, n_1\}} \{F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)\} \quad T_l'' = \psi_T(\lambda_{Me}, \lambda_a) \quad T_{MA}'' = \psi_T(\lambda_d'', \lambda_l'')$$

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^*$$

$$T_a^* = \sum_{i=1}^{n_2} X_{2i}^*$$

$$T_b^* = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i}^* > \tilde{M})$$

**Tabella 8.9** : studio via simulazione, modello (M.i.o)

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)												
$n_1 = 15, n_2 = 10; \alpha = 0.05; MC = 2000, B = 1000; \delta_2 = 0.61, 0.5, 1.168, 0.408, 0.74, 0.66;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.4185	.3575	.3615	.4170	.3415	.3970	.4010	.4040	.3990	.4185	.3120	.2815
Laplace	.3870	.4065	.3980	.3445	.4175	.4095	.3815	.3655	.4030	.3475	.3635	.2760
Cauchy	.3650	.4670	.4590	.2075	.4585	.4430	.3905	.3160	.4290	.1570	.4225	.2790
Esponenziale	.4390	.4125	.2910	.2715	.2860	.4295	.2980	.2870	.3935	.2850	.3305	.2810
Contaminate'	.3630	.3745	.3925	.2120	.3705	.3715	.3480	.2820	.3680	.0855	.3305	.2590
Half-Cauchy	.4135	.4675	.2430	.1455	.2945	.4425	.2245	.1955	.4020	.1270	.3770	.2525

H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)												
$n_1 = 15, n_2 = 10; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.0390	.0410	.0430	.0440	.0500	.0390	.0440	.0410	.0450	.0430	.0280	.0410
Laplace	.0410	.0320	.0440	.0510	.0470	.0400	.0430	.0450	.0440	.0550	.0280	.0370
Cauchy	.0510	.0510	.0490	.0510	.0480	.0580	.0500	.0410	.0540	.0300	.0400	.0420
Esponenziale	.0430	.0460	.0470	.0480	.0540	.0510	.0460	.0450	.0430	.0570	.0360	.0490
Contaminate'	.0500	.0580	.0580	.0520	.0540	.0530	.0590	.0510	.0550	.0310	.0460	.0470
Half-Cauchy	.0550	.0500	.0520	.0630	.0540	.0540	.0560	.0570	.0550	.0560	.0390	.0460

$$T_{AD}^* = \sum_{i=1}^{n_1} \frac{[F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)]}{\sqrt{[\hat{F}(X_{1(i)}^*)] \cdot [1 - \hat{F}(X_{1(i)}^*)]}}$$

$$T_d'' = \psi_T(\lambda_{AD}, \lambda_{KS})$$

$$T_{ab}'' = \psi_T(\lambda_a, \lambda_b)$$

$$T_{KS}^* = \sup_{i \in \{1, \dots, n_1\}} \{F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)\}$$

$$T_l'' = \psi_T(\lambda_{Me}, \lambda_a)$$

$$T_{MA}'' = \psi_T(\lambda_d'', \lambda_l'')$$

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^*$$

$$T_a^* = \sum_{i=1}^{n_2} X_{2i}^*$$

$$T_b^* = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i}^* > \tilde{M})$$

**Tabella 8.10** : studio via simulazione, modello (M.i.o)

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)												
$n_1 = n_2 = 15; \alpha = 0.05; MC = 2000, B = 1000; \delta_2 = 0.61, 0.5, 1.168, 0.408, 0.74, 0.66;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.4770	.4265	.4235	.4985	.4225	.4635	.4770	.4735	.4665	.4995	.3715	.3605
Laplace	.4490	.4680	.4645	.4155	.5110	.4720	.4550	.4355	.4690	.4210	.4275	.3530
Cauchy	.4380	.5495	.5455	.2060	.5965	.5175	.4600	.3935	.5070	.1485	.5060	.3585
Esponenziale	.5365	.4610	.3720	.3180	.3620	.5455	.3850	.3210	.4870	.3140	.4415	.3515
Contaminate'	.4205	.4345	.4500	.1880	.4635	.4285	.4080	.3120	.4275	.0685	.3830	.3290
Half-Cauchy	.5315	.5265	.3775	.1710	.3710	.5815	.3445	.2405	.5045	.1075	.5175	.3650

H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)												
$n_1 = n_2 = 15; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.0560	.0550	.0600	.0500	.0550	.0560	.0570	.0500	.0570	.0510	.0410	.0480
Laplace	.0470	.0500	.0500	.0510	.0660	.0480	.0450	.0440	.0480	.0500	.0390	.0450
Cauchy	.0450	.0410	.0430	.0500	.0570	.0450	.0410	.0450	.0400	.0310	.0290	.0530
Esponenziale	.0520	.0580	.0590	.0500	.0700	.0520	.0540	.0520	.0520	.0500	.0390	.0480
Contaminate'	.0550	.0540	.0530	.0510	.0800	.0570	.0530	.0550	.0490	.0240	.0450	.0560
Half-Cauchy	.0380	.0540	.0460	.0370	.0730	.0460	.0390	.0390	.0460	.0180	.0410	.0520

$$T_{AD}^* = \sum_{i=1}^{n_1} \frac{[F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)]}{\sqrt{[\hat{F}(X_{1(i)}^*)] \cdot [1 - (\hat{F}(X_{1(i)}^*))]}}$$

$$T_d'' = \psi_T(\lambda_{AD}, \lambda_{KS})$$

$$T_{ab}'' = \psi_T(\lambda_a, \lambda_b)$$

$$T_{KS}^* = \sup_{i \in \{1, \dots, n_1\}} \{F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)\}$$

$$T_l'' = \psi_T(\lambda_{Me}, \lambda_a)$$

$$T_{MA}'' = \psi_T(\lambda_d'', \lambda_l'')$$

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^*$$

$$T_a^* = \sum_{i=1}^{n_2} X_{2i}^*$$

$$T_b^* = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i}^* > \tilde{M})$$



**Tabella 8.11** : studio via simulazione, modello (M.i.o)

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)												
$n_1 = 10, n_2 = 20; \alpha = 0.05; MC = 1500, B = 750; \delta_2 = 0.61, 0.5, 1.168, 0.408, 0.74, 0.66;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}'''$	$t$	$KS^+$	$WMW$
Normale	.4420	.3773	.3980	.4560	.2333	.4187	.4307	.4307	.4207	.4527	.4173	.3080
Laplace	.3993	.4000	.4233	.3560	.2967	.4167	.4060	.4047	.4180	.3627	.4500	.3127
Cauchy	.3913	.5073	.4827	.1973	.3840	.4660	.3933	.4227	.4500	.1413	.5527	.3100
Esponenziale	.5500	.4273	.4013	.3407	.1733	.5180	.4153	.3260	.4913	.3127	.4593	.3293
Contaminate'	.4067	.4040	.4287	.1853	.2660	.4193	.3680	.3333	.4007	.0780	.4540	.3167
Half-Cauchy	.5300	.4727	.4080	.1900	.1653	.5340	.3740	.2380	.5027	.0660	.5093	.3173

H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)												
$n_1 = 10, n_2 = 20; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}'''$	$t$	$KS^+$	$WMW$
Normale	.0460	.0490	.0600	.0470	.0240	.0480	.0540	.0530	.0530	.0460	.0610	.0420
Laplace	.0410	.0460	.0390	.0410	.0240	.0380	.0410	.0430	.0410	.0440	.0600	.0420
Cauchy	.0580	.0510	.0530	.0530	.0200	.0550	.0530	.0470	.0530	.0350	.0600	.0510
Esponenziale	.0650	.0550	.0520	.0590	.0300	.0590	.0550	.0620	.0630	.0420	.0640	.0620
Contaminate'	.0480	.0410	.0460	.0430	.0310	.0440	.0470	.0510	.0470	.0380	.0520	.0420
Half-Cauchy	.0500	.0490	.0430	.0430	.0300	.0510	.0510	.0520	.0500	.0080	.0580	.0430

$$T_{AD}^* = \sum_{i=1}^{n_1} \frac{[F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)]}{\sqrt{[\hat{F}(X_{1(i)}^*)] \cdot [1 - \hat{F}(X_{1(i)}^*)]}}$$

$$T_d'' = \psi_T(\lambda_{AD}, \lambda_{KS}) \quad T_{ab}'' = \psi_T(\lambda_a, \lambda_b)$$

$$T_{KS}^* = \sup_{i \in \{1, \dots, n_1\}} \{F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)\} \quad T_l'' = \psi_T(\lambda_{Me}, \lambda_a) \quad T_{MA}''' = \psi_T(\lambda_d'', \lambda_l'')$$

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^*$$

$$T_a^* = \sum_{i=1}^{n_2} X_{2i}^*$$

$$T_b^* = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i}^* > \tilde{M})$$

**Tabella 8.12** : studio via simulazione, modello (M.i.o)

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)												
$n_1 = 20, n_2 = 10; \alpha = 0.05; MC = 1500, B = 750; \delta_2 = 0.61, 0.5, 1.168, 0.408, 0.74, 0.66;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.4380	.3747	.4060	.4560	.2367	.4113	.4413	.4313	.4280	.4613	.4093	.3053
Laplace	.4067	.4533	.4240	.3633	.3113	.4280	.4080	.4267	.4340	.3593	.4807	.3087
Cauchy	.3807	.4940	.4767	.1980	.3740	.4487	.3900	.4107	.4427	.1447	.5327	.2933
Esponenziale	.4607	.4507	.2780	.2600	.2220	.4640	.2800	.3160	.4120	.2840	.4933	.2933
Contaminate'	.3947	.3900	.4047	.1720	.2573	.3967	.3600	.3227	.3947	.0787	.4253	.2920
Half-Cauchy	.4167	.5367	.2447	.1220	.2360	.4827	.2027	.2673	.4093	.1200	.5760	.2720

H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)												
$n_1 = 20, n_2 = 10; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.0490	.0580	.0550	.0490	.0300	.0520	.0540	.0510	.0560	.0570	.0630	.0500
Laplace	.0470	.0520	.0420	.0480	.0210	.0450	.0440	.0460	.0450	.0470	.0600	.0430
Cauchy	.0450	.0410	.0380	.0480	.0180	.0440	.0400	.0430	.0410	.0350	.0480	.0460
Esponenziale	.0500	.0520	.0520	.0470	.0230	.0520	.0530	.0470	.0610	.0530	.0620	.0410
Contaminate'	.0510	.0470	.0460	.0490	.0200	.0470	.0460	.0380	.0440	.0320	.0550	.0520
Half-Cauchy	.0430	.0440	.0510	.0490	.0180	.0440	.0420	.0400	.0440	.0600	.0550	.0510

$$T_{AD}^* = \sum_{i=1}^{n_1} \frac{[F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)]}{\sqrt{[\hat{F}(X_{1(i)}^*)] \cdot [1 - \hat{F}(X_{1(i)}^*)]}}$$

$$T_d'' = \psi_T(\lambda_{AD}, \lambda_{KS})$$

$$T_{ab}'' = \psi_T(\lambda_a, \lambda_b)$$

$$T_{KS}^* = \sup_{i \in \{1, \dots, n_1\}} \{F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)\}$$

$$T_l'' = \psi_T(\lambda_{Me}, \lambda_a)$$

$$T_{MA}'' = \psi_T(\lambda_d'', \lambda_l'')$$

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^*$$

$$T_a^* = \sum_{i=1}^{n_2} X_{2i}^*$$

$$T_b^* = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i}^* > \tilde{M})$$

**Tabella 8.13** : studio via simulazione, modello (M.i.o)

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)												
$n_1 = n_2 = 20; \alpha = 0.05; MC = 1500, B = 750; \delta_2 = 0.61, 0.5, 1.168, 0.408, 0.74, 0.66;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.5720	.4680	.4987	.5800	.4513	.5413	.5480	.5360	.5420	.5773	.4440	.4333
Laplace	.5793	.5880	.6160	.5020	.6193	.5953	.5860	.5480	.6047	.5053	.5633	.4693
Cauchy	.5333	.6673	.6767	.2347	.7020	.6260	.5793	.5240	.6193	.1647	.6407	.4460
Esponenziale	.6567	.6087	.4553	.3953	.3927	.6827	.4647	.4047	.6220	.3960	.5987	.4693
Contaminate'	.5460	.5440	.5780	.1440	.5447	.5607	.5013	.3713	.5400	.0653	.5187	.4420
Half-Cauchy	.6487	.6887	.4427	.1553	.4093	.7133	.3780	.2693	.6360	.1047	.6800	.4500

H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)												
$n_1 = n_2 = 20; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0;$												
distribuzioni	test parziali di permutazione					test- aspetto		test globali		test di confronto		
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_b$	$T_d''$	$T_l''$	$T_{ab}''$	$T_{MA}''$	$t$	$KS^+$	$WMW$
Normale	.0470	.0410	.0450	.0470	.0490	.0430	.0470	.0400	.0450	.0480	.0350	.0440
Laplace	.0590	.0570	.0490	.0550	.0590	.0600	.0590	.0560	.0570	.0560	.0480	.0510
Cauchy	.0450	.0390	.0460	.0560	.0490	.0430	.0420	.0410	.0420	.0320	.0360	.0450
Esponenziale	.0680	.0530	.0590	.0650	.0580	.0670	.0680	.0550	.0740	.0670	.0440	.0470
Contaminate'	.0470	.0400	.0430	.0480	.0550	.0470	.0470	.0420	.0470	.0260	.0350	.0560
Half-Cauchy	.0550	.0550	.0640	.0600	.0620	.0540	.0640	.0570	.0580	.0320	.0490	.0480

$$T_{AD}^* = \sum_{i=1}^{n_1} \frac{[F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)]}{\sqrt{[\hat{F}(X_{1(i)}^*)] \cdot [1 - \hat{F}(X_{1(i)}^*)]}}$$

$$T_d'' = \psi_T(\lambda_{AD}, \lambda_{KS})$$

$$T_{ab}'' = \psi_T(\lambda_a, \lambda_b)$$

$$T_{KS}^* = \sup_{i \in \{1, \dots, n_1\}} \{F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)\}$$

$$T_l'' = \psi_T(\lambda_{Me}, \lambda_a)$$

$$T_{MA}'' = \psi_T(\lambda_d'', \lambda_l'')$$

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^*$$

$$T_a^* = \sum_{i=1}^{n_2} X_{2i}^*$$

$$T_b^* = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i}^* > \tilde{M})$$

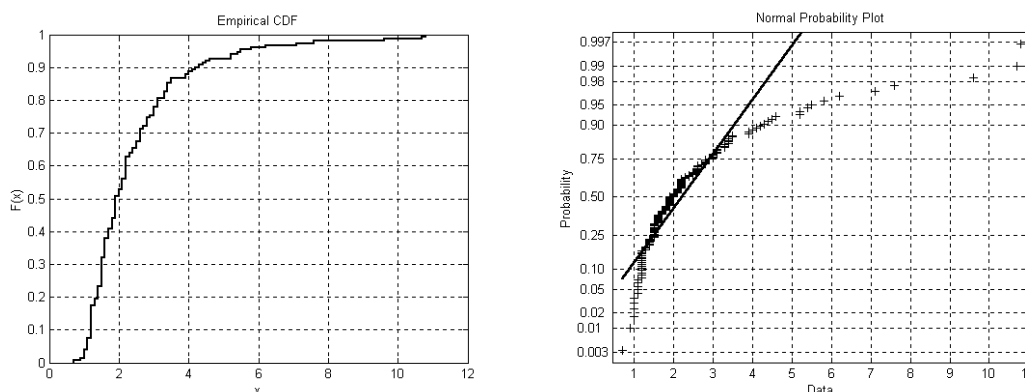
## 8.9 ANALISI DEL TEMPO PER CHIAMATE DI SERVIZIO

Si consideri un'applicazione del test bi-aspect per i dati presenti in Foster *et al.* (1998) illustrata in Marozzi (2003). Sono stati utilizzati due metodi per addestrare gli agenti di servizio di una compagnia telefonica. Viene misurato il tempo in ore per svolgere ognuna delle chiamate di servizio, in un campione complessivo di 150 unità. Le chiamate sono state suddivise casualmente in due gruppi, con 75 chiamate assegnate agli agenti addestrati con il primo metodo, e le restanti per gli agenti addestrati con il secondo metodo.

Assumiamo che il primo metodo di addestramento porta a tempi di servizio più brevi rispetto a quello degli agenti addestrati con il secondo metodo:  $H_0 : \{X_1 = X_2\}$  contro  $H_0 : \{X_1 < X_2\}$ .

Ci si può aspettare, dal momento che i dati sono tempi di attesa, che l'assunto di normalità richiesto per condurre un test  $t$  di Student venga violato. La funzione di distribuzione empirica e il *normal probability plot* mostrano che i dati sono pesantemente asimmetrici, e che certe chiamate risultano considerevolmente più lunghe di altre (Figura 8.4).

**Figura 8.4:** funzione di distribuzione empirica e *normality probability plot* per i dati in Foster *et al.* (1998)



La presenza di asimmetria e di *outliers* danneggia la prestazione del test  $t$  di Student ma non quella del test bi-aspect. Il test  $t$  di Student e l'equivalente di permutazione indicano che i due metodi di addestramento conducono a tempi di servizio paragonabili; al contrario,  $T_{ab}''$  indica che il primo metodo è migliore (si ottengono chiamate più brevi) rispetto al secondo (Tabella 8.5, prima colonna).

Abbiamo metodologicamente replicato lo studio di Marozzi per valutare le prestazioni del test globale  $T_{MA}'''$ , utilizzando lo stesso numero di permutazioni (ricampionamenti condizionati CMC), vale a dire  $B = 10000$  (Tabella 8.4):

**Tabella 8.14:** Applicazione del test  $T_{ab}''$  e  $T_{MA}'''$  per i dati in Foster *et al.* (1998)

Marozzi (2002)		test parziali e globali di permutazione				test di confronto	
<i>test</i>	<i>p-value</i>	<i>test parziale</i>	<i>p-value</i>	<i>test globale</i>	<i>p-value</i>	<i>test</i>	<i>p-value</i>
$t$	0.2583	$T_a$	0.2527	$T_{ab}''^{\geq}$	0.0110	$t[\log(X_{ji})]$	0.0414
$T_a$	0.2553	$T_b^{\geq}$	0.0061	$T_{ab}''^{>}$	0.1902	$WMW$	0.0150
$T_b$	0.0055	$T_b^{>}$	0.1209	$T_{MA}'''$	0.0031	$KS^+$	0.0021
$T_{ab}''$	0.0101	$T_{Me}$	0.0012				
		$T_{AD}$	0.0067				
		$T_{KS}$	0.0022				
		$T_l''$	0.0022				
		$T_d''$	0.0037				

*Osservazione 8.6:* Con riferimento all'Osservazione 6.7 al fine di riprodurre metodologicamente l'applicazione di Marozzi, nella tabella 8.14 sono state utilizzate le stime dei  $p$ -values in accordo alla seguente definizione di livello di significatività osservato:  $p_{obs} = \Pr\{T \geq T_{obs}; H_0\}$  quindi, in generale,  $\hat{\lambda} = \hat{L}(z | \mathbf{X}) = \{\sum_{r=1}^B \mathbf{I}(T^* \geq z)\} / B$ .

Per completezza, si è inoltre calcolato il test  $t$  di Student dopo aver effettuato una trasformazione al logaritmo dei dati ( $t[\log(X_{ji})]$ ), il test di Wilcoxon-Mann-Whitney ( $WMW$ ) e infine il test di Kolmogorov-Smirnov unilaterale ( $KS^+$ ) (tabella 8.14ultima colonna). Nella colonna centrale della tabella 8.14 vengono presentate due versioni di  $T_b$  rappresentanti rispettivamente le statistiche test  $T_b^{\geq} = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i} \geq \tilde{M})$  e  $T_b^{>} = T_b = \sum_{i=1}^{n_2} \mathbf{I}(X_{2i} > \tilde{M})$ . Di conseguenza verranno considerate due versioni della statistica test globale  $T_{ab}$ , corrispondenti alle due statistiche test parziali  $T_b^{\geq}$  e  $T_b^{>}$ . Questa scelta è dovuta al fatto che la mediana complessiva  $\tilde{M}$  risulta pari a 1.9, valore presente una volta nel primo campione e nove volte nel secondo. Ciò comporta che il valore osservato della statistica test parziale  $T_{bo}^{\geq} = 50$  mentre  $T_{bo}^{>} = 41$ . Dai risultati ottenuti la scelta della statistica test parziale utilizzata risulta determinante per l'inferenza:  $T_{ab}''^{>}$  accetta  $H_0$  mentre  $T_{ab}''^{\geq}$  la rifiuta. Una possibile soluzione potrebbe essere quella di "perturbare" i dati, ad esempio aggiustandoli sommando  $\varepsilon$  ad ogni valore, dove  $\varepsilon$  rappresenta una variabile aleatoria normale con media nulla e deviazione standard ragionevolmente piccola (ad esempio, pari a 0.00001). Tuttavia questa soluzione "artificiosa" potrebbe essere solo un modo per evitare il problema, che invece può essere

meglio compreso studiando il supporto di permutazione delle statistiche test parziali impiegate.

### 8.10 STUDIO DEL SUPPORTO DI PERMUTAZIONE DI ALCUNE STATISTICHE TEST PARZIALI IMPIEGATE

Si è considerato uno studio di simulazione per valutare la densità del supporto di permutazione delle statistiche test parziali  $T_a, T_b^{\geq}, T_b^>$ , e  $T_{Me}$ . Lo studio riprodurrà la procedura di ricampionamento senza reinserimento (CMC) per ispezionare lo spazio campionario di permutazione  $\Omega_{/X}$  e quindi definire la stima del supporto di permutazione indotto da  $T$  su  $\mathbf{X}$ :  $\hat{\tau}(\mathbf{X}) = \{T^* = T(\mathbf{X})^* : \mathbf{X}^* \in \hat{\Omega}_{/X}\} = \{T^* = T_r^*, r = 1, \dots, B\}$ . Si utilizzeranno  $B = 1000$  ricampionamenti condizionati in modo da riprodurre lo stesso grado di accuratezza utilizzato nello studio di simulazione descritto al paragrafo 8.8. Lo studio terrà in considerazione se:

- 1) le mediane  $\tilde{M}_j$  e  $\tilde{M}$  (rispettivamente del  $j$ -esimo campione e del campione complessivo) utilizzate dalle statistiche test  $T_{Me}$  e  $T_b$ , assumeranno valori presenti o meno nell'insieme dei dati  $\mathbf{X}$  a seconda che  $n_j$  o  $n$  sia pari o dispari (tabella 8.15):

**Tabella 8.15:** Valori assunti dalle mediane  $\tilde{M}_j$  e  $\tilde{M}$

$n_1$	$n_2$	$n$	$\tilde{M}_1$	$\tilde{M}_2$	$\tilde{M}$
p	p	p	$\frac{1}{2} \cdot [X_{1(n_1/2)}^* + X_{1((n_1/2)+1)}^*]$	$\frac{1}{2} \cdot [X_{2(n_2/2)}^* + X_{2((n_2/2)+1)}^*]$	$\frac{1}{2} \cdot [X_{(n/2)} + X_{((n/2)+1)}]$
p	d	d	$\frac{1}{2} \cdot [X_{1(n_1/2)}^* + X_{1((n_1/2)+1)}^*]$	$X_{2((n_2+1)/2)}^*$	$X_{((n+1)/2)}$
d	p	d	$X_{1((n_1+1)/2)}^*$	$\frac{1}{2} \cdot [X_{2(n_2/2)}^* + X_{2((n_2/2)+1)}^*]$	$X_{((n+1)/2)}$
d	d	p	$X_{1((n_1+1)/2)}^*$	$X_{2((n_2+1)/2)}^*$	$\frac{1}{2} \cdot [X_{(n/2)} + X_{((n/2)+1)}]$

- 2)  $X$  è una variabile aleatoria discreta oppure continua, cioè se la probabilità di osservare valori uguali sia rispettivamente positiva o nulla;

Lo studio di simulazione verrà strutturato secondo i seguenti criteri:

- (i)  $X_{11}, \dots, X_{1n_1}$  e  $X_{21}, \dots, X_{2n_2}$  due campioni casuali indipendenti secondo il modello con effetti additivi fissi descritto nel paragrafo 8.2.1;

- (ii)  $(n_1, n_2, n) = (10, 10, 20), (10, 11, 21), (11, 10, 21), (11, 11, 22)$ ;
- (iii)  $X$  continua  $\Rightarrow Z_{ji} \sim N(0, 1)$ ,  $X$  discreta  $\Rightarrow Z_{ji} \sim Poisson(10)$ ;
- (iv)  $B = 1000$  ricampionamenti condizionati CMC e  $MC = 2000$  simulazioni Monte Carlo; per ogni generazione di dati si considererà la densità del supporto di permutazione e l'esito per le statistiche test  $T_a, T_b^>, T_b^{\geq}$ , e  $T_{Me}$ ; infine, delle 2000 generazioni si calcolerà la densità massima, minima e media dei supporti di permutazione ottenuti per ciascuna statistica test e la proporzione di rifiuti sotto  $H_0$  (stima dell'errore di I tipo,  $\delta_2 = 0$ ) e sotto  $H_1$  (stima della potenza,  $\delta_2 = 0.61$  per  $Z_{ji} \sim N(0, 1)$  e  $\delta_2 = 2$  per  $Z_{ji} \sim Poisson(10)$ );

**Tabella 8.16:** studio del supporto di permutazione di alcune statistiche test impiegate

$n$	$n_1$	$n_2$	stat. test	$Z_{ji} \sim N(0, 1)$				$Z_{ji} \sim Poisson(10)$				
				max[ $\hat{\tau}(\mathbf{X})$ ]	min[ $\hat{\tau}(\mathbf{X})$ ]	media[ $\hat{\tau}(\mathbf{X})$ ]	$\hat{\alpha}$ $1 - \hat{\beta}$	max[ $\hat{\tau}(\mathbf{X})$ ]	min[ $\hat{\tau}(\mathbf{X})$ ]	media[ $\hat{\tau}(\mathbf{X})$ ]	$\hat{\alpha}$ $1 - \hat{\beta}$	
20	10	10	$T_a$	$H_0$	1000	994	999.01	.0530	59	18	38.58	.0595
				$H_1$	1000	993	998.89	.3010	60	18	40.20	.3560
10	10	10	$T_b^>$		9	6	6.83	.0485	9	3	6.76	.0445
					8	6	6.86	.2310	9	4	6.76	.2585
10	10	10	$T_b^{\geq}$		9	6	6.83	.0485	9	3	6.72	.0520
					8	6	6.86	.2310	9	3	6.75	.2700
10	10	10	$T_{Me}$		268	216	240.49	.0540	32	4	15.87	.0505
					262	214	240.65	.3350	32	5	16.56	.3135
21	10	11	$T_a$		1000	996	999.53	.0405	61	21	39.79	.0520
					1000	996	999.46	.4015	63	23	41.55	.4045
10	11	11	$T_b^>$		9	6	7.09	.0190	9	3	6.88	.0395
					9	6	7.08	.1935	9	3	6.87	.2335
11	11	11	$T_b^{\geq}$		9	6	7.10	.0585	9	3	6.89	.0570
					9	6	7.085	.3540	9	3	6.90	.2960
11	11	11	$T_{Me}$		168	130	150.44	.0435	30	2	15.42	.0475
					166	134	150.60	.3585	33	4	15.97	.3110
21	11	10	$T_a$		1000	994	999.50	.0560	60	21	39.85	.0580
					1000	996	999.43	.3855	68	24	41.76	.4150
11	10	10	$T_b^>$		9	6	7.08	.0570	9	3	6.87	.0515
					9	6	7.09	.3480	9	3	6.90	.2945
10	10	10	$T_b^{\geq}$		9	6	7.09	.0215	9	2	6.86	.0340
					9	6	7.09	.1805	9	3	6.87	.2560
10	10	10	$T_{Me}$		166	133	150.16	.0565	32	4	15.30	.0485
					167	136	150.32	.3295	31	2	16.06	.3300
22	11	11	$T_a$		1000	996	999.77	.0535	61	20	40.57	.0530
					1000	996	999.74	.4065	67	22	42.59	.4050
11	11	11	$T_b^>$		9	6	7.31	.0435	9	4	7.14	.0390
					9	6	7.30	.2920	9	4	7.16	.2500
11	11	11	$T_b^{\geq}$		9	6	7.31	.0435	9	5	7.12	.0385
					9	6	7.30	.2920	9	3	7.15	.2570
11	11	11	$T_{Me}$		69	53	61.96	.0515	14	1	7.64	.0450
					69	53	62.13	.3505	17	2	8.00	.2885

Nel caso di numerosità complessiva dispari ( $\Rightarrow \Pr\{\tilde{M} \in \mathbf{X}\} = 1$ ) e numerosità del secondo campione dispari,  $T_b^>$  risulta conservativo e quindi meno potente di  $T_b^{\geq}$ ; mentre con numerosità del secondo campione pari, avviene il contrario. Si può notare inoltre che il supporto di  $T_{Me}$  si riduce in densità nei casi con  $n_2$  e/o  $n_1$  dispari; questa riduzione è particolarmente evidente quando entrambi sono dispari.

## 8.11 STUDIO VIA SIMULAZIONE PER IL MODELLO CON EFFETTI STOCASTICI GENERALIZZATI

Per valutare le prestazioni dei test che verranno utilizzati, sono state condotte simulazioni Monte Carlo al fine di stimarne la potenza con il numero di proporzioni di rifiuto dell'ipotesi nulla quando essa risulta falsa, e l'errore di I tipo con il numero di proporzioni di rifiuto dell'ipotesi nulla quando essa risulta vera. Verranno effettuate MC = 1000 simulazioni Monte Carlo,  $B = 500$  permutazioni (ricampionamenti condizionati CMC), con un livello di significatività  $\alpha$  fissato al 5%.

### 8.11.1 MODELLO E IPOTESI

Siano  $X_{11}, \dots, X_{1n_1}$  e  $X_{21}, \dots, X_{2n_2}$  due campioni casuali indipendenti ottenuti da popolazioni distribuite secondo il seguente modello con effetti stocastici generalizzati (M.ii):

$$\begin{aligned} X_{1i} &= Z_{1i} \\ X_{2i} &= \Delta_{2i} + Z_{2i} + Z_{2i} = \delta_2 + W_{2i} + Z_{2i} \quad i = 1, \dots, n_j, j = 1, 2 \end{aligned}$$

dove  $Z_{ji}$  sono errori casuali scambiabili con distribuzione comune  $P$  tali che  $Z_{ji} > 0$  in probabilità,  $\Delta_{ji}$  sono effetti stocastici generalizzati rappresentanti il  $j$ -esimo livello di trattamento, che sono assunti essere uguali a zero per il primo gruppo ( $\Delta_{1i} = 0$ ) e non negativi con probabilità uno per il secondo gruppo ( $\Delta_{2i} \geq 0$ ) e che soddisfano alla condizione  $(\Delta_{2i} + Z_{2i})^2 \geq (Z_{2i})^2$ ,  $i = 1, \dots, n_2$ .

In maniera più specifica  $\Delta_{2i} = \delta_2 + W_{2i} \geq 0$ , dove  $\delta_2$  rappresenta l'effetto medio,  $W$  (errori stocastici individuali sugli effetti) quantità casuali che dipendono dagli errori  $Z$  nella seguente forma:

$$\begin{aligned} Z_{ji} &= Z_{ji1} + Z_{ji2} + Z_{ji3} \\ W_{2i} &= a \cdot Z_{2i1} + b \cdot Z_{2i4} \quad a, b \geq 0 \end{aligned}$$



dove  $Z_{jiz} \sim Q$  ( $j = 1, 2, i = 1, \dots, n_j, z = 1, \dots, 3$  (4 se  $j = 2, H_1$  e  $b \neq 0$ ), mentre  $a$  e  $b$  sono due parametri non negativi tali che:

$$(a = 0, b = 0) \Rightarrow W_{2i} = 0; \Rightarrow (\text{M.i.o})^1$$

$$(a = 0, b \neq 0) \Rightarrow W_{2i} \text{ e } Z_{2i} \text{ sono indipendenti;}$$

$$(a \neq 0, b = 0) \Rightarrow W_{2i} \text{ è dipendente da una componente di } Z_{2i};$$

$$(a \neq 0, b \neq 0) \Rightarrow W_{2i} \text{ e } Z_{2i} \text{ sono dipendenti;}$$

Si dovrà dunque verificare  $H_0 : \{X_1 = X_2\}^d$  contro  $H_1 : \{X_1 < X_2\}^d$ . E' stata considerata una configurazione con uguale numerosità campionaria:  $(n_1, n_2) = (10, 10)$ .

Per stimare il tasso di errore di I tipo abbiamo considerato  $\Delta_{2i} = 0$ , e quindi sotto  $H_0$ .

Per stimare la potenza, abbiamo considerato valori positivi di  $\Delta_{2i}$  ( $> 0$ , e quindi sotto  $H_1$ ) per i quali corrisponda approssimativamente una potenza di .25, per il test di Wilcoxon-Mann-Whitney (impiegato come *standard* di comparazione) con numerosità campionaria  $(n_1, n_2) = (10, 10)$ ,  $\alpha = .05$  e con MC = 10 000. Per essere più precisi, i valori di  $\delta_2$ ,  $a$  e  $b$  sono stati selezionati in modo tale che:

$$\delta_2 = d : \Pr\{1 - \hat{\beta}(WMW | MC) | \delta_2 = d, a = 0, b = 0; H_1\} \cong 0.125$$

$$a = a' : \Pr\{1 - \hat{\beta}(WMW | MC) | \delta_2 = d, a = a', b = 0; H_1\} \cong 0.25$$

$$b = b' : \Pr\{1 - \hat{\beta}(WMW | MC) | \delta_2 = d, a = 0, b = b'; H_1\} \cong 0.25$$

$$a = a'' : \Pr\{1 - \hat{\beta}(WMW | MC) | \delta_2 = 0, a = a'', b = 0; H_1\} \cong 0.25$$

$$b = b'' : \Pr\{1 - \hat{\beta}(WMW | MC) | \delta_2 = 0, a = 0, b = b''; H_1\} \cong 0.25$$

$$\Pr\{1 - \hat{\beta}(WMW | MC) | \delta_2 = d, a = a', b = b'; H_1\} \geq 0.25$$

$$\Pr\{1 - \hat{\beta}(WMW | MC) | \delta_2 = 0, a = a'', b = b''; H_1\} \geq 0.25$$

e si sono pertanto considerate le seguenti configurazioni:

$$(\delta_2 = 0) \cap (a = 0) \cap (b = 0) \Rightarrow \text{sotto } H_0; \Delta_{2i} = 0$$

$$(\delta_2 = d) \cap (a = a') \cap (b = 0) \Rightarrow \text{sotto } H_1; \delta_2 > 0, a > 0, b = 0$$

$$(\delta_2 = d) \cap (a = 0) \cap (b = b') \Rightarrow \text{sotto } H_1; \delta_2 > 0, a = 0, b > 0$$

$$(\delta_2 = d) \cap (a = a') \cap (b = b') \Rightarrow \text{sotto } H_1; \delta_2 > 0, a > 0, b > 0$$

$$(\delta_2 = 0) \cap (a = a'') \cap (b = 0) \Rightarrow \text{sotto } H_1; \delta_2 = 0, a > 0, b = 0$$

$$(\delta_2 = 0) \cap (a = 0) \cap (b = b'') \Rightarrow \text{sotto } H_1; \delta_2 = 0, a = 0, b > 0$$

$$(\delta_2 = 0) \cap (a = a'') \cap (b = b'') \Rightarrow \text{sotto } H_1; \delta_2 = 0, a > 0, b > 0$$

<sup>1</sup> questo caso non verrà considerato perché già discusso al paragrafo 8.5;

*Osservazione 8.7:* Questo modello è consistente con la nozione di randomizzazione, infatti: (i) nel contesto della randomizzazione, le unità sono assunte essere assegnate casualmente ai livelli di trattamento, pertanto  $H_0$  implica scambiabilità delle risposte; (ii) nell'alternativa, il trattamento può influenzare congiuntamente sia le posizioni che i coefficienti di scala, pertanto rientriamo nel contesto del problema (ristretto) di Behrens-Fisher.

### 8.11.2 STATISTICHE TEST E DISTRIBUZIONI CONSIDERATE

Verranno considerate le statistiche test di permutazione descritte nei paragrafi 8.3, 8.4, 8.5 (test parziali) e 8.6 (test globali) relative al modello (M.ii).

Come confronto, si considererà il test parametrico  $t$  di Student (unidirezionale) applicato alle osservazioni e alla loro trasformazione logaritmica<sup>1</sup> e i test nonparametrici di Wilcoxon-Mann-Whitney ( $WMW$ ) e di Kolmogorov-Smirnov ( $KS^+$ , unidirezionale).

Le distribuzioni considerate sono le seguenti (tabella 8.17):

**Tabella 8.17:** Distribuzioni considerate per il modello (M.ii)

<i>distribuzioni</i>	$\beta_1$	$\beta_2$	<i>moda</i>	<i>mediana</i>	<i>media</i>	<i>varianza</i>
<i>Uniforme</i> $Z_{jiz} \sim U(0, 8)$	0	1.8	-	4	4	5.33
<i>Chi-quadrato</i> $Z_{jiz} \sim \chi_4^2$	1.41	6	2	3.36	4	8
<i>Esponenziale</i> $Z_{jiz} \sim \exp(4)$	2	9	0	2.77	4	16
<i>Log-Normale</i> $Z_{jiz} \sim \text{Logn}(0.9,1)$	6.18	113.9	0.9	2.46	4.05	28.26
<i>Half-Cauchy</i> $Z_{jiz} \sim H-C(0,1)$	-	-	0	-	-	-

L'anormalità o curtosi è misurata dal coefficiente di eccesso o di curtosi di Pearson  $\beta_2$ , mentre l'asimmetria è misurata dall'indice  $\beta_1$  di Pearson.

*Osservazione 8.8:* Si noti che le distribuzioni considerate (esclusa quella Half-Cauchy) hanno medie uguali e sono in ordine crescente per varianza, asimmetria e curtosi, decrescente per la mediana. Inoltre, poiché  $Z_{ji} = Z_{ji1} + Z_{ji2} + Z_{ji3}$ , avremo  $Z_{ji} \sim \chi_{12}^2$ , Erlang(3,4) o  $\Gamma(3,4)$  rispettivamente per  $Z_{jiz} \sim \chi_4^2$  e  $\exp(4)$ .

<sup>1</sup> la trasformazione al logaritmo permette di ridurre il rapporto delle varianze campionarie: infatti bisogna stare attenti nell'utilizzare il test  $t$  perché c'è il rischio che il tasso di errore di I tipo superi quello nominale (Ludbrook e Dudley, 1998);

**Tabella 8.18** : studio via simulazione, modello (M.ii)

H<sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)  
 $n_1 = n_2 = 10; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 1.5, 1.6, 2.2, 2, 1.4, a' = 0.3, 0.4, 0.6, 0.65, 1.1, b' = 0;$

distr.	test parziali				test- aspetto			jls	MA	test di confronto			
	T <sub>AD</sub>	T <sub>KS</sub>	T <sub>Me</sub>	T <sub>a</sub>	T <sub>d</sub> <sup>''</sup>	T <sub>l</sub> <sup>''</sup>	T <sub>s</sub> <sup>''</sup>	T <sub>jls</sub> <sup>'''</sup>	T <sub>MA</sub> <sup>'''</sup>	t	t(log)	KS <sup>+</sup>	WMW
Unif.	.3520	.2900	.3270	.3760	.3445	.3570	.3885	.3765	.3605	.3800	.3500	.2190	.2510
Chi-q.	.3770	.3010	.3200	.3640	.3520	.3460	.3310	.3470	.3520	.3610	.3800	.2180	.2410
Espo.	.3880	.2940	.3310	.3780	.3710	.3610	.3240	.3530	.3680	.3790	.4050	.2170	.2670
Log-n.	.3970	.3020	.3250	.3290	.3810	.3460	.2840	.3160	.3580	.3060	.4000	.2170	.2830
H-Cau.	.3820	.3120	.2720	.2020	.3690	.2680	.1600	.2000	.3110	.1320	.3200	.2380	.2520

H<sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)  
 $n_1 = n_2 = 10; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 1.5, 1.6, 2.2, 2, 1.4, a' = 0, b' = 0.25, 0.3, 0.4, 0.4, 0.6;$

distr.	test parziali				test- aspetto			jls	MA	test di confronto			
	T <sub>AD</sub>	T <sub>KS</sub>	T <sub>Me</sub>	T <sub>a</sub>	T <sub>d</sub> <sup>''</sup>	T <sub>l</sub> <sup>''</sup>	T <sub>s</sub> <sup>''</sup>	T <sub>jls</sub> <sup>'''</sup>	T <sub>MA</sub> <sup>'''</sup>	t	t(log)	KS <sup>+</sup>	WMW
Unif.	.3635	.2825	.3240	.3810	.3505	.3575	.3725	.3775	.3595	.3820	.3705	.2020	.2430
Chi-q.	.3510	.2990	.3190	.3420	.3410	.3440	.3100	.3220	.3400	.3460	.3650	.2230	.2570
Espo.	.3480	.2650	.3230	.3220	.3390	.3280	.2730	.3060	.3200	.3270	.3820	.1960	.2260
Log-n.	.3890	.3110	.3170	.2840	.3610	.3220	.2300	.2730	.3370	.2730	.2610	.2200	.2470
H-Cau.	.3980	.3240	.2900	.1960	.3810	.2790	.1610	.2030	.3200	.1410	.3190	.2430	.2490

H<sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)  
 $n_1 = n_2 = 10; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 1.5, 1.6, 2.2, 2, 1.4, a' = 0.3, 0.4, 0.6, 0.65, 1.1, b' = 0.25, 0.3, 0.4, 0.4, 0.6;$

distr.	test parziali				test- aspetto			jls	MA	test di confronto			
	T <sub>AD</sub>	T <sub>KS</sub>	T <sub>Me</sub>	T <sub>a</sub>	T <sub>d</sub> <sup>''</sup>	T <sub>l</sub> <sup>''</sup>	T <sub>s</sub> <sup>''</sup>	T <sub>jls</sub> <sup>'''</sup>	T <sub>MA</sub> <sup>'''</sup>	t	t(log)	KS <sup>+</sup>	WMW
Unif.	.5630	.4640	.5090	.5920	.5540	.5645	.6055	.5965	.5745	.5970	.5700	.3690	.4375
Chi-q.	.5380	.4460	.4850	.5500	.5230	.5340	.5060	.5370	.5280	.5420	.5650	.3550	.4170
Espo.	.5220	.4150	.4760	.5040	.4980	.5020	.4400	.4830	.4950	.5140	.5700	.3260	.3890
Log-n.	.5420	.4260	.4320	.4480	.5170	.4600	.3650	.4230	.4760	.4280	.5380	.3320	.3860
H-Cau.	.5290	.4680	.4150	.3080	.5270	.4090	.2460	.3140	.4610	.2110	.4720	.3780	.3900

$$T_{AD}^* = \sum_{i=1}^{n_1} \frac{[F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)]}{\sqrt{[\hat{F}(X_{1(i)}^*)] \cdot [1 - (\hat{F}(X_{1(i)}^*))]}} \quad T_d'' = \psi_T(\lambda_{AD}, \lambda_{KS}) \quad T_{jls}^* = \psi_F(\lambda_l'', \lambda_s)$$

$$T_{KS}^* = \sup_{i \in \{1, \dots, n_1\}} \{F_1^*(X_{1(i)}^*) - \hat{F}(X_{1(i)}^*)\} \quad T_l'' = \psi_T(\lambda_{Me}, \lambda_a) \quad T_{MA}''' = \psi_T(\lambda_d'', \lambda_{jls}''')$$

$$T_{Me}^* = \tilde{M}_2^* - \tilde{M}_1^* \quad T_s'' = \sum_{i=1}^{n_2} (X_{2i}^*)^2$$

$$T_a^* = \sum_{i=1}^{n_2} X_{2i}^*$$

**Tabella 8.19** : studio via simulazione, modello (M.ii)

H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)													
$n_1 = n_2 = 10; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0, a'' = 0.75, 0.9, 1.5, 1.6, 3.1, b'' = 0;$													
distr.	test parziali				test- aspetto			jls	MA	test di confronto			
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_d''$	$T_l''$	$T_s''$	$T_{jls}''$	$T_{MA}''$	$t$	$t(\log)$	$KS^+$	WMW
Unif.	.3365	.3095	.3430	.3785	.3365	.3725	.4135	.3985	.3815	.3790	.3330	.2165	.2455
Chi-q.	.3510	.2920	.3140	.3770	.3310	.3640	.3700	.3710	.3450	.3790	.3690	.1890	.2390
Espo.	.3390	.3015	.3395	.3905	.3390	.3680	.3989	.3930	.3625	.3795	.3520	.2245	.2525
Log-n.	.3520	.2720	.3240	.3360	.3450	.3430	.3100	.3220	.3370	.3290	.3540	.1950	.2310
H-Cau.	.3520	.3050	.3030	.2730	.3420	.3070	.2430	.2700	.3190	.1680	.3350	.2240	.2400
H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)													
$n_1 = n_2 = 10; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0, a'' = 0, b'' = 0.65, 0.75, 1.1, 1.25, 2;$													
distr.	test parziali				test- aspetto			jls	MA	test di confronto			
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_d''$	$T_l''$	$T_s''$	$T_{jls}''$	$T_{MA}''$	$t$	$t(\log)$	$KS^+$	WMW
Unif.	.3470	.2740	.3150	.3650	.3390	.3415	.3600	.3565	.3465	.3580	.3535	.2055	.2420
Chi-q.	.3610	.2920	.3350	.3610	.3510	.3550	.3440	.3660	.3520	.3720	.3700	.2140	.2680
Espo.	.3420	.2840	.3130	.3410	.3310	.3410	.3170	.3300	.3410	.3480	.3730	.2080	.2490
Log-n.													
H-Cau.	.3900	.3230	.3080	.2560	.3780	.3070	.2140	.2460	.3410	.1540	.3470	.2310	.2620
H <sub>0</sub> FALSA : PROPORZIONE DI RIFIUTI (STIMA DELLA POTENZA)													
$n_1 = n_2 = 10; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0, a'' = 0.75, 0.9, 1.5, 1.6, 3.1, b'' = 0.65, 0.75, 1.1, 1.25, 2;$													
distr.	test parziali				test- aspetto			jls	MA	test di confronto			
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_d''$	$T_l''$	$T_s''$	$T_{jls}''$	$T_{MA}''$	$t$	$t(\log)$	$KS^+$	WMW
Unif.	.7420	.6790	.7335	.8055	.7500	.7905	.8310	.8120	.7925	.8105	.7560	.5950	.6640
Chi-q.	.7600	.6440	.6970	.7670	.7280	.7450	.7310	.7530	.7460	.7650	.7720	.5680	.6490
Espo.	.7370	.6250	.6870	.7520	.7220	.7390	.7240	.7400	.7400	.7530	.7580	.5320	.6250
Log-n.	.7580	.6410	.6900	.6970	.7310	.7150	.6300	.6760	.7160	.6810	.7640	.5490	.6320
H-Cau.	.7380	.6380	.6190	.4790	.7270	.6000	.4080	.4870	.6620	.3520	.6670	.5510	.5790
H <sub>0</sub> VERA : PROPORZIONE DI RIFIUTI (STIMA DELL'ERRORE DI I TIPO)													
$n_1 = n_2 = 10; \alpha = 0.05; MC = 1000, B = 500; \delta_2 = 0, a = 0, b = 0;$													
distr.	test parziali				test- aspetto			jls	MA	test di confronto			
	$T_{AD}$	$T_{KS}$	$T_{Me}$	$T_a$	$T_d''$	$T_l''$	$T_s''$	$T_{jls}''$	$T_{MA}''$	$t$	$t(\log)$	$KS^+$	WMW
Unif.	.0454	.0540	.0490	.0545	.0530	.0565	.0540	.0585	.0550	.0535	.0490	.0280	.0630
Chi-q.	.0540	.0550	.0570	.0510	.0540	.0540	.0500	.0520	.0580	.0550	.0540	.0340	.0620
Espo.	.0580	.0590	.0590	.0550	.0550	.0610	.0590	.0570	.0610	.0590	.0560	.0340	.0620
Log-n.	.0550	.0590	.0600	.0630	.0570	.0600	.0610	.0640	.0590	.0570	.0600	.0340	.0680
H-Cau.	.0500	.0470	.0550	.0570	.0450	.0560	.0630	.0560	.0550	.0330	.0480	.0290	.0390

### 8.12 CONFRONTO DELLE CONCENTRAZIONI DI COLESTEROLO NEL PLASMA DOPO AVER MANGIATO A REGIMI DI PESCE O DI CARNE

Questo esempio è ipotetico, riportato in Ludbrook e Dudley (1998). Dodici uomini sono stati reclutati tra coloro che partecipano ad una clinica di benessere o che hanno chiesto di partecipare ad un esperimento per stabilire se mangiare pesce (ma non carne) riduce la concentrazione di colesterolo nel plasma rispetto a coloro che mangiano carne (ma non pesce). I soggetti sono stati assegnati casualmente a mangiare a regime di pesce ( $n_1 = 7$ ) e di carne ( $n_2 = 5$ ). Alla termine di un anno viene misurato la loro concentrazione colesterolo nel plasma:

Mangiatori di pesce: 5.42, 5.86, 6.16, 6.55, 6.80, 7.00, 7.11

Mangiatori di carne: 6.51, 7.56, 7.61, 7.84, 11.50

Il rapporto delle varianze campionarie è 7.16 e la varianza maggiore è associata al campione più piccolo, quindi bisogna stare attenti nell'applicare il test  $t$  di Student poiché vi è il rischio che il tasso d'errore di I tipo superi quello nominale. Si potrebbero trasformare i dati al logaritmo prima di applicare il test  $t$ , e ciò ridurrebbe il rapporto delle varianze a 4.47.

Per i dati osservati si applicherà il test  $t$  di Student (unilaterale,  $t$  e  $t[\log]$  per la trasformazione dei dati al logaritmo), il test di Kolmogorov-Smirnov (unilaterale,  $KS^+$ ) e il test di Wilcoxon-Mann-Whitney ( $WMW$ )<sup>2</sup> e infine il test multi-aspect  $T_{MA}'''$  presentato nel precedente paragrafo (tabella 8.21).

**Tabella 8.20:** Risultati dei test di significatività su dati da un ipotetico esperimento che confronta le concentrazioni dei plasma colesterolo dopo aver mangiato a regimi di pesce o di carne, alternativa unilaterale. (Ludbrook e Dudley, 1998)

<i>test</i>		<i>p-values</i>		<i>p-values per la trasformazione al logaritmo</i>	
<i>t di Student</i>		.0205		.0145	
<i>Kologorov-Smirnov</i>		.0102		.0102	
<i>Wilcoxon-Mann-Whitney</i>		.0303		.0303	
<i>test parziali di permutazione</i>	<i>p-values</i>	<i>aspect e macro-aspect test</i>	<i>p-values</i>	<i>multi-aspect test</i>	<i>p-values</i>
$T_{AD}$	.0088	$T_d''$	.0038	$T_{MA}'''$	.0038
$T_{KS}$	.0101	$T_l''$	.0088		
$T_{Me}$	.0340	$T_s''$	.0038		
$T_a$	.0038	$T_{jls}'''$	.0038		

<sup>2</sup> i test di Kolmogorov-Smirnov e di Wilcoxon-Mann-Whitney sono insensibili alla trasformazione al logaritmo;

In Ludbrook e Dudley (1998) l'insieme dei dati viene analizzato per alternative bilaterali e utilizzando il test di permutazione che confronta le due medie dei campioni  $T_a$  al posto del test  $T_{MA}'''$ , e al posto del test di Kolmogorov-Smirnov viene impiegato il test di Welch (sia per i dati osservati che per la loro trasformazione al logaritmo). Ci sono 792 possibili permutazioni dei dati; il test  $T_a$  (bilaterale) ha una distribuzione di permutazione con forma bimodale e asimmetrica; sette delle possibili permutazioni che calcolano la differenza fra le medie dei due gruppi sono uguali o eccedono in una o nell'altra direzione la differenza osservata di 1.79. Ciò corrisponde ad un  $p$ -value bidirezionale pari a  $7 / 792 = .0088$  (tabella 8.21):

**Tabella 8.21:** Risultati dei test di significatività su dati da un ipotetico esperimento che confronta le concentrazioni del plasma colesterolo dopo aver mangiato a regimi di pesce o di carne, alternativa bilaterale. (Ludbrook e Dudley, 1998)

<i>test</i>	<i>p-values</i>	<i>p-values per la trasformazione al logaritmo</i>
<i>t di Student</i>	.041	.029
<i>Welch</i>	.105	.070
<i>Wilcoxon-Mann-Whitney</i>	.030	.030
$T_a$	.009	.013

# **Appendice**

### FUNZIONI DI COMBINAZIONE

```
*****
FUNZIONE DI COMBINAZIONE
*****
function [p2] = combTippett(p) %oppure% combFisher(p) %oppure% combLiptak(p)
% in ingresso:
    % p = matrice di p-values;

% in uscita:
    % il vettore di p-values ottenuti dalla funzione di combinazione;

T_sec = max(1 - p, [], 2); % funzione di combinazione di Tippett;
%oppure%
T_sec = -sum(log(p), 2); % funzione di combinazione di Fisher;
%oppure%
T_sec = sum(norminv(1-P), 2); % funzione di combinazione di Liptak;

for i = 1:size(T_sec,1)
    p2(i,1) = sum(T_sec >= T_sec(i));
end

p2 = p2./size(T_sec,1);
*****
FUNZIONI DI COMBINAZIONE
*****
function [p2] = combFLT(p) % combinazioni di Fisher, Liptak, Tippett;

[p2] = [combFisher(p) combLiptak(p) combTippett(p)];
*****
```

### SIMULAZIONI CAPITOLO 7

```
*****
SIMULAZIONE GENERALE
*****
function [results] = simGeDi(B,k,z,probz,n1,n2,MC)
% in ingresso:
    % B: numero di ricampionamenti condizionati CMC;
    % k: numero di loci (variabili);
    % z: numero di loci sotto H1;
    % probz: frequenza dell'allele raro sotto H1;
    % n1, n2: numerosità campionarie rispettivamente del 1° e del 2° campione;
    % MC: numero di simulazioni Monte Carlo;

% in uscita:
    % proporzioni di rifiuti dell'ipotesi nulla globale: stima della potenza (H0 falsa)
    % o dell'errore di I tipo (H0 vera);

alpha = 0.05; % livello di significatività assegnato;

% proporzione di rifiuti per il test globale di perm. di 3°ordine per H1 <->(bidirezionale);
pdr_tg_two = zeros(1,1);
% proporzione di rifiuti per il test globale di perm. di 3°ordine per H1 ->(unidirezionale);
pdr_tg_one = zeros(1,1);
% proporzioni di rifiuti per i test parziali di perm. di 2°ordine per H1 <->;
pdr_tp_two = zeros(1,3);
% proporzioni di rifiuti per i test parziali di perm. di 2°ordine per H1 ->;
pdr_tp_one = zeros(1,3);
% proporzione di rifiuti per il test somma di statistiche Chi-quadrato;
pdr_sChi2 = zeros(1,1);

for i = 1:MC % ciclo di simulazioni Monte Carlo: MC generazioni di dati;

    dat = randatbi(k,z,probz,n1,n2); % generazione della matrice [n x k] di dati;

    [p_sChi2] = sumChi2(dat,k,n1,n2); % p-value per la somma di stat. Chi-quadrato;
    rif_sChi2 = (p_sChi2 <= alpha); % funzione indicatrice del rifiuto di H0;
    pdr_sChi2 = pdr_sChi2 + rif_sChi2; % aggiornamento della prop. di rifiuti;
```



appendice

```
[p_two,p_one] = TpermGeDi(dat,n1,n2,B); % matrici di p-values per i k test parziali

[mp2_two] = combFLT(p_two); % matr. di p per comb. di Fisher,Liptak,Tippett per H1 <->;
[mp2_one] = combFLT(p_one); % matr. di p per comb. di Fisher,Liptak,Tippett per H1 ->;

[p2_two] = mp2_two(B+1,:); % p-values di 2° ord. <-> con comb.Fisher,Liptak,Tippett;
rif_tp_two = (p2_two <= repmat(alpha,1,3)); % funzione indicatrice del rifiuto di H0;
pdr_tp_two = pdr_tp_two + rif_tp_two; % aggiornamento delle prop. di rifiuti;
[p2_one] = mp2_one(B+1,:); % p-values di 2° ord. -> con comb. Fisher,Liptak,Tippett;
rif_tp_one = (p2_one <= repmat(alpha,1,3)); % funzione indicatrice del rifiuto di H0;
pdr_tp_one = pdr_tp_one + rif_tp_one; % aggiornamento delle prop. di rifiuti;

mp3_two = combT(mp2_two); % matr. di p di 3° ord. <-> con comb. Tippett;
p3_two = mp3_two(B+1); % p-value per il test globale <->;
mp3_one = combT(mp2_one); % matr. di p di 3° ord. -> con comb. Tippett;
p3_one = mp3_one(B+1); % p-value per il test globale ->;

rif_tg_two = (p3_two <= alpha); % funzione indicatrice del rifiuto di H0;
pdr_tg_two = pdr_tg_two + rif_tg_two; % aggiornamento delle prop. di rifiuti;
rif_tg_one = (p3_one <= alpha); % funzione indicatrice del rifiuto di H0;
pdr_tg_one = pdr_tg_one + rif_tg_one; % aggiornamento delle prop. di rifiuti;
end

% si relativizza la prop. di rif. con rispetto a MC;
pdr_tg_two = pdr_tg_two/MC;
pdr_tg_one = pdr_tg_one/MC;
pdr_tp_two = pdr_tp_two/MC;
pdr_tp_one = pdr_tp_one/MC;
pdr_sChi2 = pdr_schi2/MC;

[results] = [pdr_tg_two pdr_tg_one pdr_tp_two pdr_tp_one pdr_sChi2];
*****
SUMMATION METHOD
*****
function [p_sChi2] = sumChi2(dat,k,n1,n2)
% in ingresso:
% dat: matrice di dati [n x k];
% k: numero di loci (variabili);
% n1, n2: numerosità campionarie rispettivamente del 1° e del 2° campione;

% in uscita:
% p-value della statistica somma di statistiche Chi-quadrato;
%(summation method applicato alla statistica Chi-quadrato tradizionale);

n = n1 + n2; % numerosità complessiva;
sumChi2 = 0;

for i=1:k % ciclo che calcola le k distinte statistiche Chi-quadrato e le somma;

piuno = sum(dat(1:n1,i));
piudue = sum(dat(n1+1:n,i));
zeruno = n1-piuno;
zerdue = n2-piudue;
[i_cont] = [piuno piudue; zeruno zerdue]; % i-esima tabella di contingenza 2 x 2;

vett_chi = chi22(i_cont); % valore della stat. Chi-quadrato per l'i-esima tab. di cont.;

sumChi2 = sumChi2 + vett_chi; % aggiornamento della statistica somma di stat. Chi-quadr.;

end

p_sChi2 = (1-chi2cdf(chidue,k)); %p-value della statistica somma di stat. Chi-quadrato;
*****
TEST CHI-QUADRATO PER TABELLE DI CONTINGENZA 2x2
*****
function vett_chi22 = chi22(i_cont)
% in ingresso:
% i_cont: tabella di contingenza 2x2;
% in uscita:
% vett_chi22: valore della statistica Chi-quadrato;
```

```

sizem = size(i_cont);
margx = sum(i_cont,1);
margy = sum(i_cont,2);
n = sum(margx);

matexp = zeros(sizem); % matrice 2x2 attesa;
for i = 1:2
    for j = 1:2
        matexp(j,i,:) = margx(1,i,:).*margy(j,1,:)./n;
    end;
end;

vett_chi22 = sum(sum((i_cont-matexp).^2./matexp));
*****
TEST PARZIALI DI PERMUTAZIONE
*****
function [p_two, p_one] = TpermGeDi(dat,n1,n2,B)
% in ingresso:
    % dat: matrice di dati [n x k];
    % n1,n2: numerosità campionarie rispettivamente del 1° e del 2° campione;
    % B = numero di ricampionamenti condizionati CMC;

% in uscita:
    % matrice di p-values [(B+1) x k] (p_one) per il test unidirezionale;
    % matrice di p-values [(B+1) x k] (p_two) per il test bidirezionale;

[n k] = size(dat);
n = n1+n2;
counter = zeros(B+1,k);

for i = 1:B % ciclo delle B permutazioni casuali dei dati;
    perm = randperm(n);
    Tuno(i,:) = (sum( dat(perm(1:n1),:)))/n1;
    Tdue(i,:) = (sum( dat(perm(n1+1:n),:)))/n2;
    Tperm_one(i,:) = Tdue(i,:) - Tuno(i,:); % Test T*h ->;
end

Tuno(B+1,:) = (sum( dat((1:n1),:)))/n1;
Tdue(B+1,:) = (sum( dat((n1+1:n),:)))/n2;
Tperm_one(B+1,:) = Tdue(B+1,:)-Tuno(B+1,:); % Test T*h_oss ->;

Tperm_two = abs(Tperm_one); % Test T*h <->;

for i = 1:B+1 % ciclo per calcolare le matrici di p_values;
    for j = 1:k
        counter_two(i,j) = sum(Tperm_two(i,j)<=Tperm_two(:,j));
        counter_one(i,j) = sum(Tperm_one(i,j)<=Tperm_one(:,j));
    end
end

p_two = [counter_two/(B+1)]; % matrice di p-values [(B+1) x k] (p_one) per il test ->;
p_one = [counter_one/(B+1)]; % matrice di p-values [(B+1) x k] (p_two) per il test <->;
*****

```

## SIMULAZIONI CAPITOLO 8

```

*****
SIMULAZIONE GENERALE (M.io)
*****
function [results] = simMio%nomedistr%(B,delta2,n1,n2,MC) % modello di risposta (M.i.o);

% in ingresso:
    % B: numero di ricampionamenti condizionati CMC;
    % delta2: parametro di spostamento;
    % n1, n2: numerosità campionarie rispettivamente del 1° e del 2° campione;
    % MC: numero di simulazioni Monte Carlo;
    % nomedistr: distribuzione dalla quale si generano i dati:
    %     Normale(normrnd), t di Student(trnd), Uniforme(unifrnd),
    %     Esponenziale(exprnd), Chi-quadrato n.c.(ncx2rnd); per le distribuzioni
    %     Laplace, Bimodale, Half-Cauchy e Contaminate si veda Osservazione 8.5;

```

appendice

```

% in uscita:
    % proporzioni di rifiuti dell'ipotesi nulla globale: stima della potenza (H0 falsa)
    % o dell'errore di I tipo (H0 vera);

alpha = .05; % livello di significatività assegnato;

pdr_tds = zeros(1,1); % proporzione di rifiuti per il test t di Student (unidirezionale);
pdr_wmw = zeros(1,1); % proporzione di rifiuti per il test Wilcoxon-Mann-Whitney;
pdr_ks2 = zeros(1,1); % proporzione di rifiuti per il test Kolmogorov-Smirnov(unidirez.);
pdr_Tab = zeros(1,1); % proporzione di rifiuti per il test di perm. globale Tab;
pdr_Tl = zeros(1,1); % proporzione di rifiuti per il test di perm. aspetto Tl;
pdr_Td = zeros(1,1); % proporzione di rifiuti per il test di perm. aspetto Td;
pdr_TMA = zeros(1,1); % proporzione di rifiuti per il test di perm. globale TMA;
pdr_Tp = zeros(1,5); % proporzioni di rif. per i test parz. di perm. TAD,TKS,Tme,Ta,Tb;

n = n1 + n2;
for i = 1:MC % ciclo di simulazioni Monte Carlo: MC generazioni di dati;

    % generazione della matrice [n x 1] di dati;
    dat = %nomedistr.%(delta2,n1,n2);

    % test di confronto;
    [tds] = ttest2(dat(1:n1),dat(n1+1:n),alpha,-1); %t di Student (unilaterale);
    [p_w,h_w] = ranksumadj(dat(1:n1),dat(n1+1:n),alpha); %Wilcoxon-Mann-Whitney;
    [ks2] = kstest2(dat(1:n1),dat(n1+1:n),alpha,1); %Kolmogorov-Smirnov (unilaterale);

    % matrice di p-values per i test parziali di permutazione TAD, TKS, TMe, Ta, Tb;
    [P] = TpermMio(dat,n1,n2,B);

    % combinazioni dei test di permutazione;
    [mp_Tab] = combTippett(P(:,4:5)); % test di perm. globale T"ab;
    [p_Tab] = mp_Tab(B+1);

    [mp_Tl] = combTippett(P(:,3:4)); % test di perm. aspetto T"l;
    [p_Tl] = mp_Tl(B+1);

    [mp_Td] = combTippett(P(:,1:2)); % test di perm. aspetto T"d;
    [p_Td] = mp_Td(B+1);

    mp_MA(:,1) = mp_Tl;
    mp_MA(:,2) = mp_Td;

    mp_TMA = combT(mp_MA); % test di perm. globale TMA;
    p_TMA = mp_TMA(B+1);

    % aggiornamenti delle proporzioni di rifiuto;
    pdr_tds = pdr_tds + tds;
    pdr_wmw = pdr_wmw + h_w;
    pdr_ks2 = pdr_ks2 + ks2;

    rif_Tab = (p_Tab <= alpha);
    pdr_Tab = pdr_Tab + rif_Tab;

    rif_Tl = (p_Tl <= alpha);
    pdr_Tl = pdr_Tl + rifTl;

    rif_Td = (p_Td <= alpha);
    pdr_Td = pdr_Td + rif_Td;

    rif_TMA = (p_TMA <= alpha);
    pdr_TMA = pdrTMA + rif_TMA;

    [p_Tp] = P(B+1,:);
    rif_Tp = (p_Tp <= repmat(alpha,1,5));
    pdr_Tp = pdr_Tp + rif_Tp;
end

% si relativizza la prop. di rifiuti con rispetto a MC;
pdr_tds = pdr_tds/MC;
pdr_wmw = pdr_wmw/MC;
pdr_ks2 = pdr_ks2/MC;
pdr_Tab = pdr_Tab/MC;

```

appendice

```

pdr_Tl = pdr_Tl/MC;
pdr_Td = pdr_Td/MC;
pdr_TMA = pdr_TMA/MC;
pdr_Tp = pdr_Tp/MC;

results=[pdr_Tp pdr_Td pdr_Tl pdr_MA pdr_tds pdr_ks2 pdr_wmw];
*****
SIMULAZIONE GENERALE (M.ii)
*****
function [results] = simMii%nomedistr%(B,delta2,a,b,n1,n2,MC) % modello di risposta (M.ii);
% in ingresso:
% B: numero di ricampionamenti condizionati CMC;
% delta2: parametro di spostamento;
% a, b: parametri per la dipendenza;
% n1, n2: numerosità campionarie rispettivamente del 1° e del 2° campione;
% MC: numero di simulazioni Monte Carlo;
% nomedistr: distribuzione dalla quale si generano i dati:
%          si veda il paragrafo 8.11; Uniforme(unifrnd), Chi-quadrato (chi2rnd),
%          Esponenziale (exprnd), Log-Normale(lognrnd); Half-Cauchy (Osserv. 8.5);

% in uscita:
% proporzioni di rifiuti dell'ipotesi nulla globale: stima della potenza (H0 falsa)
% o dell'errore di I tipo (H0 vera);

alpha = .05; % livello di significatività assegnato;

pdr_tds = zeros(1,1); % proporzione di rifiuti per il test t di Student (unidirezionale);
pdr_wmw = zeros(1,1); % proporzione di rifiuti per il test Wilcoxon-Mann-Whitney;
pdr_ks2 = zeros(1,1); % proporzione di rifiuti per il test Kolmogorov-Smirnov(unidirez.);
pdr_ltds= zeros(1,1); % prop. di rif. per il test t di Student applicato al log(dat);
pdr_Tl = zeros(1,1); % proporzione di rifiuti per il test di perm. aspetto Tl;
pdr_Td = zeros(1,1); % proporzione di rifiuti per il test di perm. aspetto Td;
pdr_Ts = zeros(1,1); % proporzione di rifiuti per il test di perm. aspetto Ts;
pdr_Tjls= zeros(1,1); % proporzione di rifiuti per il test di perm. macroaspetto Tjls;
pdr_TMA = zeros(1,1); % proporzione di rifiuti per il test di perm. globale TMA;
pdr_Tp = zeros(1,5); % proporzioni di rif. per i test parz. di perm. TAD,TKS,Tme,Ta;

n = n1 + n2;
for i = 1:MC % ciclo di simulazioni Monte Carlo: MC generazioni di dati;

% generazione della matrice [n x 1] di dati;
dat = %nomedistr.%(delta2,n1,n2,a,b); % (M.ii)

% test di confronto;
[tds] = ttest2(dat(1:n1),dat(n1+1:n),alpha,-1); %t di Student (unilaterale);
[p_w,h_w] = ranksumadj(dat(1:n1),dat(n1+1:n),alpha); %Wilcoxon-Mann-Whitney;
[ks2] = kstest2(dat(1:n1),dat(n1+1:n),alpha,1); %Kolmogorov-Smirnov (unil.);
[ltds] = ttest2(log(dat(1:n1)),log(dat(n1+1:n)),alpha,-1); %t di Stud. per log(dat);

% matrice di p-values per i test parziali di permutazione TAD, TKS, TMe, Ta;
[P] = TpermMii(dat,n1,n2,B);

% combinazioni dei test di permutazione;

[mp_Tl] = combTippett(P(:,3:4)); % test di perm. aspetto T"l;
[p_Tl] = mp_Tl(B+1);

[mp_Td] = combTippett(P(:,1:2)); % test di perm. aspetto T"d;
[p_Td] = mp_Td(B+1);

[mp_Ts] = P(:,5);
[p_Ts] = mp_Ts(B+1);

mp_JLS(:,1) = mp_Tl; % test di perm. aspetto Ts;
mp_JLS(:,2) = mp_Ts;

mp_Tjls = combF(mp_JLS); % test di perm. macro-aspetto Tjls;
p_Tjls = mp_Tjls(B+1);

mp_MA(:,1) = mp_Tl;
mp_MA(:,2) = mp_Tjls;

mp_TMA = combT(mp_MA); % test di perm. globale TMA;

```

```

p_TMA = mp_TMA(B+1);
% aggiornamenti delle proporzioni di rifiuto;
pdr_tds = pdr_tds + tds;
pdr_wmw = pdr_wmw + h_w;
pdr_ks2 = pdr_ks2 + ks2;
pdr_ltds= pdr_ltds+ ltds;

rif_Ts = (p_Ts <= alpha);
pdr_Ts = pdr_Ts + rif_Ts;

rif_Tl = (p_Tl <= alpha);
pdr_Tl = pdr_Tl + rif_Tl;

rif_Td = (p_Td <= alpha);
pdr_Td = pdr_Td + rif_Td;

rif_Tjls = (p_Tjls <= alpha);
pdr_Tjls = pdr_Tjls + rif_Tjls;

rif_TMA = (p_TMA <= alpha);
pdr_TMA = pdr_TMA + rif_TMA;

[p_Tp] = P(B+1,:);
rif_Tp = (p_Tp <= repmat(alpha,1,4));
pdr_Tp = pdr_Tp + rif_Tp;
end

% si relativizza la prop. di rifiuti con rispetto a MC;
pdr_tds = pdr_tds/MC;
pdr_wmw = pdr_wmw/MC;
pdr_ks2 = pdr_ks2/MC;
pdr_ltds= pdr_ltds/MC;
pdr_Ts = pdr_Ts/MC;
pdr_Tl = pdr_Tl/MC;
pdr_Td = pdr_Td/MC;
pdr_Tjls= pdr_Tjls/MC;
pdr_TMA = pdr_TMA/MC;
pdr_Tp = pdr_Tp/MC;

results = [pdr_Tp pdr_Td pdr_Tl pdr_Ts pdr_Tjls pdr_TMA pdr_tds pdr_ltds pdr_ks2 pdr_wmw];
*****
TEST DI PERMUTAZIONE PARZIALE TAD
*****
function [TAD] = anddar(dat,n1,n2)
% in entrata: matrice di dati [n x 1], numerosità campionaria n1, n2;
% in uscita: valore della statistica test parziale di perm. TAD;
n = n1 + n2;

for i = 1:n
    ind_all(i) = sum(dat(i,1) >= dat(:,1));
end
dot = dat(1:n1);
for j=1:n1
    ind_c1(j) = sum(dot(j,1) >= dot(:,1));
end

temp_all = [(1:n) ./ n];
temp_c1 = [(1:n1) ./ n1];
cdf_all = temp_all(ind_all);
cdf_c1 = temp_c1(ind_c1);

for k = 1:n1
    if cdf_all(1,k) == 1
        den(1,k) = cdf_all(1,k);
    else den(1,k) = sqrt(cdf_all(1,k).*(1 - cdf_all(1,k)));
    end
end
TAD = sum((cdf_c1-cdf_all(1:n1))./den);
*****
TEST DI PERMUTAZIONE PARZIALE TKS
*****
function [TKS] = kolsmi(dat,n1,n2)
% in entrata: matrice di dati [n x 1], numerosità campionaria n1, n2;

```

appendice

```

% in uscita: valore della statistica test parziale di perm. TKS;
n = n1 + n2;

for i=1:n
    ind_all(i) = sum(dat(i,1) >= dat(:,1));
end
dot = dat(1:n1);
for j = 1:n1
    ind_c1(j) = sum(dot(j,1)>=dot(:,1));
end

temp_all = [(1:n) ./ n];
temp_c1 = [(1:n1) ./ n1];
cdf_all = temp_all(ind_all);
cdf_c1 = temp_c1(ind_c1);

Tmatr = cdf_c1 - cdf_all(1:n1);
TKS = max(Tmatr);
*****
TEST PARZIALI DI PERMUTAZIONE
*****
function [P] = TpermMio(dat,n1,n2,B); %oppure% Tpermii(dat,n1,n2,B);
% in entrata:
    % dat: matrice di dati [n x 1]
    % B: numero di ricampionamenti condizionati CMC;
    % n1, n2: numerosità campionarie rispettivamente del 1° e del 2° campione;
% in uscita:
    % (M.i.o) valori delle statistiche test parziali di permutazione TAD,TKS,TMe,Ta,Tb;
    % (M.ii) valori delle statistiche test parziali di permutazione TAD,TKS,TMe,Ta,;

n= n1 + n2;
counter = zeros(B+1,5); %oppure% zeros(B+1,4);

for i = 1:B % ciclo delle B permutazioni casuali dei dati;
    perm = randperm(n);
    Tperm(i,1) = anddar(dat(perm),n1,n2); % TAD;
    Tperm(i,2) = kolsmi(dat(perm),n1,n2); % TKS;
    Tuno(i) = median(dat(perm(1:n1)));
    Tdue(i) = median(dat(perm(n1+1:n)));
    Tperm(i,3) = Tdue(i) - Tuno(i); % TMe;
    Tperm(i,4) = sum(dat(perm(n1+1:n),1)); % Ta;
    Tperm(i,5) = sum(dat(perm(n1+1:n))>median(dat)); % Tb; non per (M.ii);
end

% valori osservati delle statistiche test parziali di permutazione;
Tperm(B+1,1)=anddar(dat,n1,n2);
Tperm(B+1,2)=kolsmi(dat,n1,n2);
Tperm(B+1,3)=median(dat(n1+1:n))-median(dat(1:n1));
Tperm(B+1,4)=sum(dat(n1+1:n),1);
Tperm(B+1,5)=sum(dat(n1+1:n)>median(dat)); non per (M.ii);

for i=1:B+1
    counter(i,1) = sum(Tperm(i,1)<=Tperm(:,1));
    counter(i,2) = sum(Tperm(i,2)<=Tperm(:,2));
    counter(i,3) = sum(Tperm(i,3)<=Tperm(:,3));
    counter(i,4) = sum(Tperm(i,4)<=Tperm(:,4));
    counter(i,5) = sum(Tperm(i,5)<=Tperm(:,5)); non per (M.ii);
end

P=[counter/(B+1)];
*****

```

# Bibliografia

- Arboretti, R., Pesarin, F., Romero, M., Salmaso, L. (2000) Il progetto SETIG e la valutazione comparativa delle strategie terapeutiche adottate: Metodologia statistica e applicazione. *Giornale italiano di Farmacia Clinica*, 14, 26-36.
- Berger, V.W. (2000) Pros and cons of permutation tests in clinical trials. *Statistics in Medicine*, 19, 1319-1328.
- Celant, G., Pesarin, F. (2000a) Alcune osservazioni critiche riguardanti l'analisi bayesiana condizionata. *Statistica*, LX, 25-37.
- Celant, G., Pesarin, F. (2000b) Sulla definizione di analisi condizionata. *Statistica*, 2, 185-194.
- Cochran, W. G. (1965) The planning of observational studies of human populations (with Discussion). *Journal of the Royal Statistical Society, Series A*, 128, 234-255.
- Conover, W. (1999) *Practical Nonparametric Statistics*. Wiley, New York.
- Cox, D.R. (1977) The role of significance tests. *Scandinavian Journal of Statistics*. 4, 49-70.
- Cox, D.R., Hinkley, D.V. (1974) *Theoretical Statistics*. Chapman & Hall, London.
- D'Agostino, R.B., Stephens, M.A. (1986) *Goodness of fit techniques*. Marcel Dekker, New York.
- Dancelli, L. (1993) Test di adattamento per distribuzioni di cui siano note frequenze e quantità. *Statistica*, 1, 87-108
- De Finetti, B. (1960) Evoluzione verso una sintesi. *Probabilità e induzione (1992), Supplemento al no. 3 di Statistica*.
- De Finetti, B. (1977) Il ruolo della probabilità nei diversi atteggiamenti del pensiero scientifico. *Probabilità e induzione (1992), Supplemento al no. 3 di Statistica*.
- Di Castelnuovo, A., Mazzaro, D., Pesarin, F., Salmaso, L. (2001) Test di permutazione multidimensionali in problemi di inferenza isotonica: un'applicazione alla genetica. *Statistica*, anno LX, no. 4.
- Diggle, P.J. (1985) *Statistical Analysis of Spatial Point Patterns*. Academic Press, London.
- Edgington, E. S., (1995) *Randomization Tests*. 3<sup>rd</sup> ed., Marcel Dekker, New York.

- Finos, L., Pesarin, F., Salmaso, L. (2000a) Test combinati per il controllo della molteplicità mediante procedure di closed testing. *Working Paper no. 2000.8, Department of Statistics, University of Padova.*
- Fisher, R. A. (1934) *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh.
- Fisher, R. A. (1935) *The Design of Experiments*. Oliver & Boyd, Edinburgh.
- Foster, D.P., Waterman, R.P. (1998) *Basic Business Statistics*. Springer-Verlag, New York.
- Foster, P.L., Eisenstadt, E., Cairns, J. (1996) Testing differentiation in diploid populations. *Genetics*, 144, 1933-1940.
- Good, P. (2000) *Permutation Tests*. 2<sup>nd</sup> Edition, Springer-Verlag, New York.
- Hollander, M. and Wolfe, D.A. (1999) *Nonparametric Statistical Methods*. 2<sup>nd</sup> Edition, Wiley, New York.
- Jorde, P.E., Ryman, N. (1996) Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change of allele frequencies. *Genetics*, 143, 1369-1381.
- Keller-McNulty S., Higgins, I. J. (1987) Effect of tail weight and outliers on power and type-I error of robust permutation tests for location. *Communications in Statistics – Computation and Simulation*, 16, 17-35.
- Kempthorne, O. (1955) The randomization theory of experimental inference. *Journal of the American Statistical Association*, 50, 964-967.
- Kempthorne, O. (1979) Sampling inference, experimental inference and observation inference. *Sankhyā*, 40, 115-145.
- Landenna, G., Marasini, D. (1990) *Metodi Statistici Non Parametrici*. Il Mulino, Bologna.
- Ludbrook, J., Dudley, H. (1998) Why permutation tests are superior to *t* and *F* tests in biomedical research. *The American Statistician*, 52, 127-132.
- Lehmann, E. L. (1986) *Testing Statistical Hypotheses* (2<sup>nd</sup> edn). Wiley, New York.
- Marozzi, M. (2002d) A Bi-Aspect nonparametric test for the two-sample location problem, *Computational Statistics and Data Analysis*. In corso di pubblicazione.
- Pesarin, F. (1986) Considerazioni sull'impostazione decisionale dell'inferenza statistica. *Statistica*, anno XLVI, no. 1, 1986.
- Pesarin, F. (1989) The analysis of spatial point patterns combining a set of indicators. In *Atti del III Congresso della Società Italiana di Ecologia*, 805-808. SIE, Siena.



- Pesarin, F. (1989) Casualità e test statistici di casualità nell'analisi crittografica. *Proc. SPRI 1989*, 168-187, Fondazione U. Bordoni, Roma.
- Pesarin, F. (2001) *Multivariate permutation tests with applications in biostatistics*. Wiley, Chichester.
- Pesarin, F. (2002) Extending permutation inferences to unconditional ones. *Statistical Methods & Applications*, 11, 161-173.
- Rosenbaum, P.R. (1984) Conditional permutation tests and the propensity score in observational studies. *Journal of American Statistical Association*, 79, 565-574.
- Rosenbaum, P.R., Rubin, D.B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Ryman, N., Jorde, P.E. (2001) Statistical power when testing for genetic differentiation. *Molecular Ecology*, 10, 2361-2373.
- Sasieni, P.D. (1997) From genotypes to genes: Doubling the sample size. *Biometrics*, 53, 1253-1261.
- Senn, S. (1994) Fisher's game with the Devil. *Statistics in Medicine*, 13, 217-230;
- Westfall, P.H., Young, S.S. (1993) *Resampling-Based Multiple Testing*. Wiley, New York.
- Westfall, P.H., Russel, D.W. (2000) Closed Multiple Testing Procedures and PROC MULTTEST. *Observations, the technical journal for SAS Software users*.