

ALCUNI PROBLEMI DI VERIFICA D'IPOTESI PER VARIABILI CATEGORIALI

Fortunato Pesarin

1. INTRODUZIONE E MOTIVAZIONI

Quando Italo Scardovi mi fece sapere della sua iniziativa, ossia di invitare alcuni colleghi che avevano conosciuto personalmente Amato Herzel e Alighiero Naddeo a scrivere degli articoli per ricordarne la figura scientifica, aderii subito. In seguito emerse il problema sulla scelta del tema da trattare. Tema che allo stesso tempo contenesse risultati recenti delle proprie ricerche e che contribuisse al ricordo di entrambi, magari collegandosi da un lato a qualcuno dei loro argomenti di studio, dall'altro a stimoli emersi in occasione di personali incontri e discussioni.

Amato Herzel, tra le tante altre cose, si era interessato a lungo e con profondità al problema della quantificazione per variabili categoriali, avendo tra gli obiettivi anche quello di costruire una metodologia generale per l'analisi multivariata di variabili categoriali ordinate. Problema notoriamente difficile quanto importante. Nell'ambito di questa problematica, ricordo la discussione seguita ad un suo seminario sull'argomento in cui aveva segnalato come molto interessante, e perciò meritevole di una adeguata soluzione sia per gli aspetti prettamente metodologici, sia per quelli di tipo computazionale e sia per le moltissime potenziali applicazioni, il problema dell'estensione a variabili categoriali di un qualche cosa di analogo del test T^2 di Hotelling. Ossia della verifica d'ipotesi dell'uguaglianza in distribuzione di due (o più) popolazioni quando vi siano implicate $Q > 1$ variabili categoriali tra loro dipendenti, sia ordinate, sia nominali e sia miste, ovvero alcune quantitative ed altre categoriali.

Alighiero Naddeo, da specialista qual era, si era interessato in particolare alla metodologia generale della verifica delle ipotesi multivariate. Di lui ricordo le molte stimolanti discussioni, principalmente a Padova, Venezia e Trento, ma anche a Roma, sia sui principi dell'inferenza statistica e sia su molti problemi di metodologia per i quali la tradizionale impostazione parametrica si mostrava in difficoltà. Tra questi in particolare, collegato al tradizionale problema delle tabelle di contingenza $2 \times C$, ricordo che aveva segnalato come piuttosto interessante quanto importante il caso per cui, qualora si dovesse pervenire al rigetto dell'ipotesi nulla di uguaglianza distributiva, ci si proponga di individuare a quale o a quali classi o modalità que-

sta significatività debba essere maggiormente attribuita. Ossia una sorta di estensione della metodologia del tipo delle comparazioni multiple, ovvero dei test multipli, per variabili categoriali o categorizzate con connesso controllo degli errori inferenziali. Ricordo che allora, forse anche per la consapevolezza della grande difficoltà che la cosa avrebbe comportato, non si era parlato del caso multivariato.

Questi problemi, assieme ad altri riconducibili alle per me più frequenti discussioni avute anche con Odoardo Cucconi, mio maestro e loro estimatore ed amico, e riconosciuti come altrettanto utili quanto intriganti per la loro complessità, mi hanno per così dire tenuto a lungo impegnato. Di tanto in tanto, dopo aver condotto tentativi di soluzione che a prima vista sembravano promettenti ma che non reggevano in modo soddisfacente all'analisi critica, lasciavo perdere per riprenderli, magari dopo un certo tempo, nella convinzione che prima o poi qualcosa di accettabile sarebbe alla fine scaturito. Una soddisfacente soluzione mi pare sia emersa, dopo innumerevoli fallimenti e delusioni, a seguito della introduzione della metodologia della combinazione non parametrica di test dipendenti. Metodologia che, una volta messa a punto e corredata dei principali teoremi di consistenza, non distorsione, ottimalità asintotica, ecc. ha dato il via alla soluzione di moltissimi problemi, tra cui anche quelli evocati. Siccome tali soluzioni sono venute per così dire quasi a cascata e senza soluzione di continuità, come gli anelli di una catena in cui uno segue l'altro, piuttosto che redigere un articolo per ognuna da pubblicare su rivista, mi è parso il caso di scrivere un libro (Pesarin, 2001) che insieme alla teoria ne raccogliesse la maggior parte. Libro che però non contiene che in parte i dettagli delle soluzioni che illustrerò nel seguito.

La sezione 2 è dedicata al problema dell'uguaglianza in distribuzione per variabili categoriali uni- e multi- variate, con anche un sommario del metodo di combinazione non parametrica di una pluralità di test dipendenti. La sezione 3 tratta il problema dell'uguaglianza in distribuzione di due popolazioni sotto il profilo dei test multipli.

2. VERIFICA D'IPOTESI PER VARIABILI CATEGORIALI ORDINATE

2.1. *Il caso unidimensionale*

In primo luogo, consideriamo in questa sezione un problema di uguaglianza in distribuzione di due popolazioni quando viene osservata una sola variabile categoriale ordinata. L'analogo problema con una variabile nominale è da ritenersi universalmente noto alla comunità degli statistici, almeno per quel che concerne le soluzioni standard (p. es. Agresti, 2002). Non così qualora nell'analisi vi sia coinvolta una pluralità di tali variabili (Pesarin, 1993, 1994, 2001). Questo problema sarà trattato nella sezione 2.3, come semplice estensione di quello univariato, tramite la metodologia della combinazione non parametrica.

Intendiamo sviluppare con un certo dettaglio l'analisi per variabili ordinate in quanto queste rivestono uno speciale interesse, sia perché si incontrano molto spesso nelle applicazioni alle scienze sociali, alla psicologia, alle prove cliniche, alle

indagini sulle preferenze di prodotti industriali e/o commerciali, alle indagini sulla qualità di servizi, e così via, e sia perché danno luogo a delicati problemi sia di impostazione matematica, sia di ordine computazionale e sia anche di tipo strettamente metodologico, molti dei quali si presentano come estremamente difficili se affrontati con le metodologie parametriche standard. Tra i molti problemi inferenziali ad esse collegati rivestono una riconosciuta importanza quelli cosiddetti di dominanza stocastica. Tra questi, emblematico è quello connesso agli esperimenti Caso/Controllo, consistente nello stabilire statisticamente se la popolazione dei Casi, ossia dei soggetti cui è stato somministrato simbolicamente il trattamento a livello 1, è stocasticamente dominata da (o domina) quella dei Controlli, ossia dei soggetti cui è stato somministrato il livello 0.

Assumiamo che il supporto di una data variabile categoriale univariata X sia ripartito in $C \geq 2$ classi o modalità ordinate $\{A_c, c = 1, \dots, C\}$, per le quali abbia senso una relazione d'ordine tra le classi del tipo $A_1 < A_2 < \dots < A_C$, come ad esempio accade con gli aggettivi “scarso” $<$ “discreto” $<$ “buono”, ecc. Supponiamo, inoltre, che i dati siano rilevati indipendentemente tra loro su un insieme di n soggetti e che siano classificati secondo due livelli di un generico trattamento. Pertanto, dati due campioni indipendenti di osservazioni $\mathbf{X}_j = \{X_{jr}, r = 1, \dots, n_j\}, j = 1, 2$, si voglia saggiare l'ipotesi nulla $H_0 : \{X_1 \stackrel{d}{=} X_2\} \equiv \{F_1(A_c) = F_2(A_c), c = 1, \dots, C-1\}$ contro l'alternativa $H_1 : \{X_1 \stackrel{d}{>} X_2\} \equiv \{\bigcup_c [F_1(A_c) < F_2(A_c)]\}$, in cui la funzione $F_j(A_c) = Pr\{X_j \leq A_c\}, j = 1, 2$, gioca il ruolo di funzione di ripartizione (cumulativa) della variabile categoriale X_j . Si noti che H_1 definisce la dominanza stocastica di X_1 rispetto a X_2 , in quanto X_1 manifesta “valori” tendenzialmente più grandi di quelli manifestati da X_2 . Inoltre, è importante osservare che H_1 nella forma $\{\bigcup_c [F_1(A_c) < F_2(A_c)]\}$ definisce un insieme di alternative cosiddette ristrette, ossia composto da $C-1$ sotto-alternative unilaterali o direzionali. In questa forma il problema si presenta come notoriamente difficile si vedano ad esempio i lavori di Hirotsu (1982, 1986, 2002), Sampson and Whitaker (1989), Dykstra *et al.* (1995), El Barmi and Dykstra (1995), Wang (1996), Dardanoni and Forcina (1998), Cohen *et al.* (2000, 2003), Perlman and Wu (2002). Si noti anche che H_0 afferma che i dati dei due gruppi sono tra loro scambiabili, e che quindi per la relativa analisi risulta applicabile il principio di permutazione.

I dati d'osservazione sono usualmente organizzati in una tabella di contingenza del tipo $2 \times C$, ossia $\{f_{jc} = \#(X_{jr} \in A_c), r = 1, \dots, n_j; c = 1, \dots, C; j = 1, 2\}$, in cui $\#(\cdot)$ indica il numero di valori osservati che soddisfano (\cdot) . I simboli $N_{jc} = \sum_{i \leq c} f_{ji}$, $f_{.c} = f_{1c} + f_{2c}, n_j = \sum_c f_{jc} = N_{jC}$, $n = n_1 + n_2$ indicano rispettivamente le frequenze cumulate e quelle marginali di riga e colonna. La tabella 1 ne riporta una tipica rappresentazione.

L'insieme costituito dalla riunione di tutti i dati (insieme ottenuto dal concatenamento delle due liste) $\mathbf{X} = \{X_{jr}, r = 1, \dots, n_j, j = 1, 2\}$ e l'insieme delle frequenze marginali $\{n_1, n_2; f_{.1}, \dots, f_{.C}\}$ sono insiemi tra loro equivalenti di statistiche sufficienti sotto H_0 in quanto, a meno di un inessenziale ordinamento all'interno delle classi, risultano legati da una relazione biunivoca.

Una soluzione ottima quando $C = 2$ è costituita dal ben noto test delle probabilità esatte di Fisher. Il quale, tra l'altro, costituisce uno dei pochissimi casi di test di permutazione la cui distribuzione è nota in forma chiusa; in questo caso Ipergeometrica.

TABELLA 1
Una tipica tabella $2 \times C$

A_c	f_{1c}	f_{2c}	$f_{\cdot c}$	N_{1c}	N_{2c}
A_1	f_{11}	f_{21}	$f_{\cdot 1}$	N_{11}	N_{21}
...
A_c	f_{1c}	f_{2c}	$f_{\cdot c}$	N_{1c}	N_{2c}
...
A_C	f_{1C}	f_{2C}	$f_{\cdot C}$	N_{1C}	N_{2C}
	n_1	n_2	n	-	-

L'analisi secondo il principio di permutazione di variabili discrete, specialmente in situazioni multidimensionali, viene molto facilitato se, in luogo delle usuali tabelle di contingenza, i dati vengono rappresentati unità per unità, ossia elencando gli n dati individuali in analogia con quanto è abituale con i dati quantitativi. Anche perché utile ai fini della trattazione via permutazione, indichiamo una tale rappresentazione per il caso di due campioni con la notazione $\mathbf{X} = \{X(r), r = 1, \dots, n; n_1, n_2\}$, dove è inteso che le prime n_1 risposte tra le n elencate sono quelle provenienti dal primo gruppo e le altre n_2 dal secondo.

Una soluzione al problema di uguaglianza in distribuzione è data dal test di permutazione

$$T_D^* = \sum_{c=1}^{C-1} (N_{2c}^* - N_{1c}^*) \left[4 \frac{N_{\cdot c}}{n} \left(\frac{n - N_{\cdot c}}{n} \right) \frac{n_1 n_2}{n-1} \right]^{-\frac{1}{2}}$$

in cui: $N_{\cdot c} = N_{1c} + N_{2c} = N_{1c}^* + N_{2c}^*$, N_{1c}^* e N_{2c}^* , $c = 1, \dots, C-1$, sono le frequenze cumulate in una generica permutazione dei dati. Permutazioni che possono essere convenientemente ottenute prendendo una permutazione (u_1^*, \dots, u_n^*) dei primi n interi $(1, \dots, n)$ e considerando il vettore permutato dei dati $\mathbf{X}^* = \{X(u_r^*), r = 1, \dots, n; n_1, n_2\}$ con l'associata tabella di contingenza $\{f_{jc}^* = \#(X_{jb}^* \in A_c), b = 1, \dots, n_j; c = 1, \dots, C; j = 1, 2\}$, in cui ovviamente: $X_{1r}^* = X(u_r^*)$ se $1 \leq r \leq n_1$, mentre $X_{2,r-n_1}^* = X(u_r^*)$ se $n_1 + 1 \leq r \leq n$. Va notato che nella corrispondente tabella permutata, i totali marginali rimangono invariati rispetto a quella osservata, ossia $f_{\cdot c}^* = f_{\cdot c}$ e $N_{\cdot c}^* = N_{\cdot c}$, $c = 1, \dots, C$, per tutte le possibili permutazioni.

La statistica test T_D è essenzialmente basata sul confronto tra le due distribuzioni empiriche (EDF) $\hat{F}_j^*(A_c) = N_{jc}^* / n_j$, $j = 1, 2$, quindi corrisponde alla versione discreta della statistica per la bontà di adattamento per alternative di domi-

nanza secondo Cramér-von Mises con aggiustamento secondo Anderson-Darling (D'Agostino and Stephens, 1986).

Di interesse sono anche le soluzioni di Dykstra *et al.* (1995), El Barmi and Dykstra (1995), Dardanoni and Forcina (1998) e Cohen *et al.* (2000, 2003) per alternative ristrette basate sul comportamento asintotico del test del rapporto di massima verosimiglianza, che dà luogo al cosiddetto “chi-quadrato barrato”, $\bar{\chi}^2$. Così come di interesse è la soluzione proposta da Hirotsu (1982, 1986, 2002), basata sul comportamento del cosiddetto “cumulative chi-squared”. E' noto che queste soluzioni presentano alcuni inconvenienti anche severi. Anzitutto, secondo Wang (1996) le distribuzioni asintotiche del rapporto di massima verosimiglianza risultano dipendenti dai parametri ignoti $\theta_c = \Pr\{X \in A_c\}$, $c = 1, \dots, C$, che peraltro risultano consistentemente stimati dalle frequenze relative marginali $\hat{\theta}_c = f_c/n$. Inoltre, non sono ancora state proposte estensioni multivariate. D'altro canto, l'uso del “cumulative chi-squared” proposto da Hirotsu, trattandosi comunque di una statistica di forma quadratica, non consente propriamente la verifica d'ipotesi di dominanza, bensì di quella di diversità, in cui l'alternativa è del tipo $H_1 : \{X_1 \neq X_2\}$. E tale soluzione, a detta dell'autore, non può essere estesa a più di due variabili.

In ogni caso, va sottolineato che con variabili categoriali nominali hanno senso soltanto analisi di diversità distributiva, poiché in tale contesto quelli di dominanza stocastica sono in generale privi di senso.

2.2. Soluzioni mediante combinazione non parametrica

Si osservi, in particolare, che il c -esimo termine di T_D^* , ossia

$$(N_{2c}^* - N_{1c}^*)[N_{\cdot c}(n - N_{\cdot c})]^{-\frac{1}{2}} = T_c^*,$$

il quale, a meno della costante moltiplicativa $[4n_1n_2/n^2(n-1)]^{\frac{1}{2}}$, irrilevante in quanto comune a tutte le permutazioni, altro non è che la forma standardizzata di una statistica che compara le due frequenze cumulate della classe c -esima in uno schema ipergeometrico per saggiare le sotto-ipotesi: $H_{0c} : \{F_c = G_c\}$, contro $H_{1c} : \{F_c < G_c\}$, corrispondenti alla c -esima sotto-tabella 2×2 , ossia $\{(N_{jc}^*, n_j - N_{jc}^*); j = 1, 2\}$, estratta dalla tabella completa $2 \times C$ (vedasi anche la sezione 3.1). Per cui si ha: $H_0 = \bigcap_c H_{0c}$ e $H_1 = \bigcup_c H_{1c}$, forma che mostra la decomposizione delle ipotesi in $C - 1$ sotto-ipotesi o ipotesi parziali.

Quindi, T_D^* può essere visto (Pesarin, 1992, 2001) come una combinazione non parametrica diretta (vedansi le sotto-sezioni che seguono) di $C - 1$ test parziali, tutti espressi in forma standardizzata, ossia $T_D^* = \sum_{c \leq C-1} T_c^*$.

Naturalmente, in una prima fase ad ognuna delle $C - 1$ sotto-tabelle estratte potrebbe essere applicato un qualunque test T marginalmente non distorto, come ad esempio quello di Fisher delle probabilità esatte, o quello del rapporto di massima verosimiglianza, o quelli basati sulle “power-divergence statistics” di Cressie and Read (1988), ecc. E, nella seconda fase, procedendo alla combinazione non parametrica dei loro rispettivi valori- p (i cosiddetti p -values) $\lambda_c = \Pr\{T_c^* \geq T_{oc} \mid \mathbf{X}\}$, in cui T_{oc} rappresenta il valore osservato del test T per la c -esima sotto-ipotesi. Si noti che la distribuzione di probabilità che viene presa in considerazione nel calcolo dei valori λ_c è quella di permutazione, cioè quella risultante dal processo di condizionamento ai dati osservati che, a loro volta, sono statistiche sufficienti in H_0 per la sottostante distribuzione di popolazione, qualunque essa sia. Questo modo di procedere è discusso in Pesarin (2001, Cap. VI). Qui lo riassumiamo brevemente per metterne in rilievo gli aspetti principali.

2.2.1 *Combinazione non parametrica di k test di permutazione dipendenti*

Se i $k > 1$ test parziali fossero tra loro stocasticamente indipendenti, il problema della loro combinazione troverebbe facili soluzioni (p. es. Folks, 1984). Il problema della combinazione si pone invece in modo piuttosto intrigante quando i test sono tra loro dipendenti e la forma della dipendenza è non nota, o difficile da trattare analiticamente o anche soltanto non lineare. In tali casi diventa necessario disporre di una metodologia generale di combinazione non parametrica, specialmente con riguardo alla sottostante struttura di dipendenza. Una via, piuttosto battuta in letteratura, è stata quella di pervenire ad un controllo approssimato del rischio d’errore di prima specie tramite disuguaglianze come quella di Bonferroni. Una via la cui performance notoriamente peggiora all’aumentare del numero di test parziali fino a renderla del tutto inutilizzabile proprio nelle situazioni in cui più servirebbe (p. es. Hochberg and Tamhane, 1987; Westfall and Young, 1993, Hirotsu *et al.*, 2001). La combinazione, con controllo esatto del rischio di prima specie, può però essere effettuata in generale nell’ambito dei metodi di permutazione, e ciò sotto condizioni molto blande e facilmente verificabili.

Per unitarietà di trattazione, la combinazione non parametrica prende in considerazione i valori- p , λ_i , associati ai test parziali di permutazione T_i , $i = 1, \dots, k$. Test parziali che sono del tutto generali ma che comunque vanno assunti come appropriati per le rispettive sotto-ipotesi. Altresì è da notare che qualcuno dei test parziali potrebbe riferirsi a sotto-alternative di dominanza ed altri a sotto-alternative di diversità, così come alcuni potrebbero riguardare variabili quantitative ed altri variabili categoriali.

In quanto segue, con riguardo ai test parziali assumiamo che:

- siano significativi per valori grandi;
- siano marginalmente (separatamente) consistenti e non distorti.

La non distorsione marginale dei test parziali, ossia che ogni T_i sotto H_{1i} assuma valori stocasticamente non più piccoli che sotto H_{0i} , comporta che i valori- p asso-

ciati $(\lambda_1, \dots, \lambda_k)$ sono positivamente dipendenti ("positively upper orthant dependent", Lehmann, 1986).

Il test combinato assume quindi la forma: $T_{\psi}'' = \psi(\lambda_1, \dots, \lambda_k)$, in cui la funzione di combinazione, continua e non degenera, $\psi : (0,1)^k \rightarrow R^1$, deve soddisfare alle condizioni:

1. È non crescente in ciascun argomento: $\psi(\dots, \lambda_i, \dots) \geq \psi(\dots, \lambda'_i, \dots)$ se $\lambda_i < \lambda'_i$, $\forall i \in (1, \dots, k)$.
 2. Raggiunge il suo limite superiore $\bar{\psi}$, che potrebbe anche essere infinito, se almeno uno degli argomenti raggiunge lo zero: $\psi(\dots, \lambda_i, \dots) \rightarrow \bar{\psi}$ se $\lambda_i \rightarrow 0$, $\forall i \in (1, \dots, k)$.
 3. Il suo valore critico $T_{\psi\alpha}''$ è finito, ossia $T_{\psi\alpha}'' < \bar{\psi}$, $\forall \alpha > 0$.
- Sotto queste condizioni, che in generale sono di facile verifica, il test combinato T_{ψ}'' risulta non distorto e consistente (Pesarin, 2001).

2.2.2 Combinazione non parametrica via simulazione Monte Carlo condizionale

La trattazione matematica in forma chiusa della distribuzione di permutazione k -variata è notoriamente impossibile, sia per la generalmente elevata cardinalità dell'insieme delle possibili permutazioni ma soprattutto per la difficoltà nella caratterizzazione analitica della regione critica del test che si voglia adottare. Risulta invece possibile e computazionalmente appropriata l'applicazione di un metodo di combinazione non parametrica dei test parziali, cui resta associata una regione critica più facilmente caratterizzabile ed il cui valore- p globale viene di fatto stimato tramite un procedimento di simulazione Monte Carlo condizionale consistente nel considerare un campione casuale semplice di B elementi dall'insieme di tutte le possibili permutazioni dei dati. Va da sé che qualora fosse possibile sondare esaurivamente l'intero spazio delle permutazioni, anziché a stime, si perverrebbe al calcolo esatto. L'algoritmo di simulazione, per un problema a due campioni, è definito dai seguenti passi:

- a) Calcolare i valori osservati delle statistiche T_i , ossia $T_{0i} = T_i(\mathbf{X})$, $i = 1, \dots, k$; valori compendati nel vettore \mathbf{T}_0 .
- b) Sia (u_1^*, \dots, u_n^*) una permutazione casuale di $(1, \dots, n)$, allora $\mathbf{X}^* = \{X(u_i^*), i = 1, \dots, n; n_1, n_2\}$ è la permutazione casuale di \mathbf{X} associata a (u_1^*, \dots, u_n^*) . Calcolare quindi i corrispondenti valori delle statistiche T_i , ossia $T_i^* = T_i(\mathbf{X}^*)$, $i = 1, \dots, k$ scritti nel vettore \mathbf{T}^* (si noti che la variabile X qui implicata è del tutto generale: uni- o multi- variata; categoriale, quantitativa o mista).
- c) Ripetere indipendentemente B volte il passo (b); quindi l'insieme dei relativi

risultati, $\{\mathbf{T}_r^*, r=1, \dots, B\}$, costituisce un campione casuale che simula la distribuzione di permutazione k -variata di \mathbf{T} .

d) L'associata funzione di ripartizione empirica

$$\hat{F}_B(\mathbf{t}) = \left[1/2 + \sum_{r=1}^B \mathbf{I}(\mathbf{T}_r^* \leq \mathbf{t}) \right] / (B+1), \mathbf{t} \in \mathbb{R}^k,$$

dove $\mathbf{I}(\cdot) = 1$ se la relazione (\cdot) è soddisfatta e 0 altrimenti, per il ben noto teorema di Glivenko-Cantelli, fornisce una stima fortemente consistente uniformemente della distribuzione multivariata $F(\mathbf{t}|\mathbf{X})$ di $\mathbf{T} = \{T_1, \dots, T_k\}$. Inoltre le funzioni

$$\hat{L}_i(t) = \left[1/2 + \sum_{r=1}^B \mathbf{I}(T_{ir}^* \geq t) \right] / (B+1), t \in \mathbb{R}^1, i=1, \dots, k,$$

stimano consistentemente i livelli di significatività marginali di T_i , per cui

$\hat{\lambda}_i = \hat{L}_i(T_{0i})$ sono le stime dei valori- p di T_i .

e) Il valore osservato del test combinato tramite la funzione di combinazione ψ è dato da $T_{\psi_0}'' = \psi(\hat{\lambda}_1, \dots, \hat{\lambda}_k)$.

f) La distribuzione di permutazione simulata del test combinato pertanto diventa

$$\{T_{\psi_r}''^* = \psi[\hat{L}_1(T_{1r}^*), \dots, \hat{L}_k(T_{kr}^*)], r=1, \dots, B\}.$$

g) Il valore- p stimato del test combinato è quindi $\hat{\lambda}_{\psi}'' = \sum_r \mathbf{I}(T_{\psi_r}''^* \geq T_{\psi_0}'') / B$.

Da notare che la stima $\hat{\lambda}_{\psi}''$ converge fortemente al vero valore $\lambda_{\psi}'' = \Pr\{T_{\psi}''^* \geq T_{\psi_0}'' | \mathbf{X}\}$ quando B tende all'infinito o quando viene esaurito l'intero spazio delle permutazioni.

h) Se $\hat{\lambda}'' \leq \alpha$, allora l'ipotesi nulla H_0 viene rigettata al livello di significatività α .

Tra le molte funzioni di combinazione non parametrica, a parte quella diretta che, per il problema di dominanza a due campioni con dati categoriali ordinati in cui i test parziali sono tutti dello stesso tipo ed espressi in forma standardizzata, dà luogo a $T_D'' = T_D = \sum_{i \leq k} T_i$, possiamo suggerire quella di Fisher, quella di Tippett o quella di Liptak, che corrispondono rispettivamente a: $T_F'' = -\sum_i \log(\lambda_i)$, $T_T'' = \max_i(1 - \lambda_i)$ e $T_L'' = \sum_i \Phi^{-1}(1 - \lambda_i)$, in cui Φ rappresenta la funzione di ripartizione della distribuzione normale standardizzata.

In particolare, è da notare che la funzione di combinazione di Tippett corrisponde a qualcosa di analogo ad una forma di statistica di Kolmogorov-Smirnov con aggiustamento secondo Anderson-Darling per variabili discrete, ovvero ad una forma come $T_{KS}^* = \max_i \{(N_{2i}^* - N_{1i}^*)[N_{\cdot i}(n - N_{\cdot i})]^{-1/2}\}$, a sua volta corrispondente ad una forma di massimo test parziale standardizzato: $T_{KS}^* = \max_i(T_i^*)$.

Per il loro buon comportamento, tanto asintotico quanto per numerosità finite e sia in termini di potenza, tra le varie soluzioni prospettate e prospettabili la preferenza di chi scrive va a T_D'' , T_F'' e T_T'' . La combinazione di Tippett viene raccomandata quando si ha sentore che una o poche sub-alternative siano attive; quella di Liptak o quella diretta quando si pensa che tutte o quasi le sub-alternative siano attive con valori delle rispettive non centralità pressoché uguali; quella di Fisher negli altri casi o qualora non si abbiano precise conoscenze su quante, e con quale noncentralità, sono le sub-alternative attive.

Sottolineiamo che tra le prerogative della metodologia di combinazione non parametrica, oltre a fornire un controllo esatto del livello di significatività per l'ipotesi globale vi è anche quella di poter pervenire ad un controllo della significatività sulle singole sub-ipotesi. Quest'ultimo aspetto sarà trattato nella sezione 3.

2.3. Alcune estensioni uni- e multi- variate

1) In un problema a due campioni con alternative non ristrette, o di non dominanza, con variabili categoriali ordinate le ipotesi sono $H_0: \{X_1 \stackrel{d}{=} X_2\}$ contro $H_1: \{X_1 \stackrel{d}{\neq} X_2\} \equiv \{\bigcup_c [F_1(A_c) \stackrel{d}{\neq} F_2(A_c)]\}$. Quindi l'estensione della statistica test T_D diventa

$$T_D^{*2} = \sum_{c=1}^{C-1} (N_{2c}^* - N_{1c}^*)^2 [N_{\cdot c} (n - N_{\cdot c})]^{-1},$$

che corrisponde alla versione per una variabile discreta del test di Cramér-von Mises-Anderson-Darling per il confronto delle distribuzioni di due campioni. Da notare che $T_D^{*2} \neq (T_D^*)^2$.

Va notato che con variabili nominali il problema della diversità distributiva è usualmente risolto con la statistica chi-quadrato (χ^2), o con statistiche desunte dal rapporto di massima verosimiglianza o con statistiche del tipo delle "power divergence" di Cressie & Read (1988). È altresì interessante notare che tali statistiche sono comparate con la loro distribuzione asintotica, la quale se le numerosità, anche entro le classi, non sono basse generalmente approssima bene quella effettiva. Distribuzione effettiva che, poiché tali test presuppongono tutti la stima di parametri di disturbo tramite statistiche sufficienti, altro non è che la distribuzione di permutazione.

2) L'estensione delle ipotesi e della statistica test a $K > 2$ campioni per alternative non ristrette e con variabili categoriali ordinate diventa ormai ovvia. Denotando con $\hat{F}_{jc}^* = N_{jc}^* / n_j$ e $\hat{F}_{\cdot c}^* = N_{\cdot c}^* / n$, in cui $N_{\cdot c} = \sum_j N_{jc} = \sum_j N_{jc}^*$, rispettivamente la distribuzione empirica parziale e marginale, chiaramente abbiamo

$$T_{KD}^{*2} = \sum_{j=1}^K \sum_{c=1}^{C-1} (\hat{F}_{jc}^* - \hat{F}_{\cdot c})^2 [\hat{F}_{\cdot c} \cdot (1 - \hat{F}_{\cdot c}) \cdot (n - n_j) / n_j]^{-1}.$$

Notiamo che, dal punto di vista della sua strutturazione, T_{KD}^{*2} corrisponde ad una forma di combinazione diretta di test parziali ottenuti comparando la distribuzione di ciascun gruppo con quella marginale, ossia per $1 \leq j \leq K$, $T_{KDj}^{*2} = \sum_c (\hat{F}_{jc}^* - \hat{F}_{\cdot c})^2 [\hat{F}_{\cdot c} \cdot (1 - \hat{F}_{\cdot c}) \cdot (n - n_j) / n_j]^{-1}$. Naturalmente, in luogo della combinazione diretta è possibile pensare ad una qualunque altra forma di combinazione non parametrica dei valori- p .

3) Per una estensione al caso multivariato in un problema di alternative di dominanza stocastica a due campioni, assumiamo che la variabile risposta categoriale ordinata sia Q -dimensionale $\mathbf{X} = (X_1, \dots, X_Q)$ in cui il numero di classi ordinate, non necessariamente costante per le varie componenti X_q , è rappresentato dal vettore $\mathbf{C} = (C_1, \dots, C_Q)$ e che le n_1 e n_2 unità statistiche di rilevazione siano osservate indipendentemente l'una dall'altra rispettivamente dalle variabili di popolazione \mathbf{X}_1 e \mathbf{X}_2 . Le ipotesi che intendiamo saggiare sono $H_0: \{\mathbf{X}_1 \stackrel{d}{=} \mathbf{X}_2\} \equiv \{\bigcap_{q=1}^Q \bigcap_{c=1}^{C_q-1} (F_{qc} = G_{qc})\}$ contro $H_1: \{\mathbf{X}_1 \stackrel{d}{>} \mathbf{X}_2\} \equiv \{\bigcup_{q=1}^Q \bigcup_{c=1}^{C_q-1} (F_{qc} < G_{qc})\}$, dove F_{qc} e G_{qc} giocano il ruolo di funzione di ripartizione marginale per la q -esima variabile componente X_q .

Questo problema (Wang, 1996) è considerato come estremamente difficile, se non impossibile, qualora lo si voglia risolvere tramite il rapporto di massima verosimiglianza, mentre nell'ambito della combinazione non parametrica di test di permutazione dipendenti la soluzione diventa particolarmente semplice. Per esempio, la combinazione diretta di Q test parziali (T_{D1}, \dots, T_{DQ}) diviene semplicemente

$$T_{MD}^{**} = \sum_{q=1}^Q \sum_{c=1}^{C_q-1} (N_{q2c}^* - N_{q1c}^*) [N_{q\cdot c} (n - N_{q\cdot c})]^{-1/2} = \sum_q T_{Dq}^*,$$

dove: $N_{qjc}^* = \sum_{r \leq c} f_{qjr}^*$ e $N_{q\cdot c} = N_{q1c} + N_{q2c}$. Ovviamente, in un problema di non dominanza a due campioni, in cui l'alternativa è $H_1: \{\mathbf{X}_1 \stackrel{d}{\neq} \mathbf{X}_2\} \equiv \{\bigcup_{q=1}^Q \bigcup_{c=1}^{C_q-1} (F_{qc} \neq G_{qc})\}$, la soluzione diventa

$$T_{MD}^{**2} = \sum_{q=1}^Q \sum_{c=1}^{C_q-1} (N_{q2c}^* - N_{q1c}^*)^2 [N_{q\cdot c} (n - N_{q\cdot c})]^{-1}.$$

La soluzione T_{MD}^{*2} corrisponde all'analogo del test T^2 di Hotelling per variabili categoriali ordinate.

4) Se le Q variabili dovessero essere tutte nominali, avrebbero senso soltanto alternative di diversità in distribuzione, ovvero di non dominanza. In questo caso, i test parziali potrebbero avere per ciascuna variabile componente X_q , ad esempio, la forma χ^2 cui resta associato il valore- p di permutazione λ_q^* , per cui la soluzione diventerebbe $T_{MF}'' = -\sum_q \log(\lambda_q^*)$, $T_{MT}'' = \max_q(1 - \lambda_q^*)$ e $T_{ML}'' = \sum_q \Phi^{-1}(1 - \lambda_q^*)$, rispettivamente per la combinazione di Fisher, di Tippett e di Liptak. Queste soluzioni corrispondono all'analogo del test T^2 di Hotelling per variabili categoriali nominali.

Qui va subito notato che una soluzione di combinazione diretta mediante la somma dei χ_q^2 parziali non è accettabile a meno che il numero di classi C_q non sia costante per tutte le Q variabili componenti. Ad esempio, in un problema con due variabili è del tutto evidente che se una avesse 10 classi e l'altra 2, per cui per la prima il χ^2 avrebbe 9 gradi di libertà mentre quello della seconda ne avrebbe 1 soltanto, e ne venisse considerata la loro semplice somma, ne conseguirebbe che il primo addendo dominerebbe stocasticamente sul secondo. Pertanto, ne risulterebbe un'inferenza sostanzialmente sbilanciata in cui di fatto la prima componente verrebbe a pesare 9 volte più della seconda.

Lasciamo al lettore per esercizio le ormai ovvie estensioni al caso di K campioni multidimensionali.

2.4. Un esempio con una variabile binaria bivariata

I dati che seguono, raccolti da Abbate et Al. (2001) e analizzati in Pesarin (2001), riguardano 65 lavoratori di una sede di una società petrolchimica sui quali è stata valutata la presenza o l'assenza di ansietà (variabile X) e di depressione (variabile Y), secondo gli items del protocollo di Hamilton. I soggetti, lavoratori che sono occupati nella medesima sede e nel medesimo reparto da un minimo di 12 a un massimo di 15 anni, sono ripartiti in due gruppi, a seconda delle condizioni ambientali del reparto in cui lavorano, uno costituito da 48 individui ad alta esposizione all'azione degli idrocarburi ed uno di 17 individui a bassa esposizione, corrispondenti a due livelli di un simbolico trattamento. I dati sono riportati nella tabella 2 che segue:

TABELLA 2
Ansietà e Depressione secondo Hamilton

$X \setminus Y$	I		II	
	0	1	0	1
0	27	1	15	0
1	13	7	2	0

le cui associate tabelle marginali sono

X	I	II	
0	28	15	43
1	20	2	22
	48	17	65

e

Y	I	II	
0	40	17	57
1	8	0	8
	48	17	65

I soggetti del primo gruppo di 48 lavoratori sono sospettati di essere esposti a condizioni che possono indurre ansietà e/o depressione in misura maggiore che non per il secondo gruppo di lavoratori, esposti invece a condizioni del tutto normali in un normale ambiente di lavoro. Quindi il secondo gruppo gioca il ruolo di gruppo di controllo. Va anche osservato, che all'atto dell'assunzione i lavoratori selezionati erano stati assegnati pressoché a caso ad uno dei due reparti, in relazione soltanto alla vacanza di posti, essendo le attività da svolgere sostanzialmente di natura molto simile. Questo giustifica l'assunto per cui, qualora le differenti condizioni di esposizione non dovessero comportare differenti effetti sulle due variabili risposta, le osservazioni bivariate sarebbero scambiabili. Pertanto, l'ipotesi nulla può essere scritta nella forma $H_0 : \{(X_1, Y_1) \stackrel{d}{=} (X_2, Y_2)\}$. Mentre l'ipotesi alternativa, dovendo interpretare la situazione in cui vi siano differenti maggiori effetti su almeno una delle variabili risposta da parte degli individui del primo reparto, può essere scritta nella forma $H_1 : \{(X_1, Y_1) \stackrel{d}{>} (X_2, Y_2)\} \equiv \{(X_1 \stackrel{d}{>} X_2) \cup (Y_1 \stackrel{d}{>} Y_2)\}$, ossia nella forma di dominanza stocastica bivariata.

Siccome le variabili risposta sono di tipo binario: $X, Y = \{1, 0\}$, a seconda che l'ansietà o la depressione siano presenti o meno, il problema corrisponde ad una estensione bivariata del test di Fisher delle probabilità esatte. Usando il software NPC Test 2.0 con $B = 10\,000$ iterazioni Monte Carlo, il test parziale per la variabile X ha dato luogo ad una stima del valore- p di permutazione $\hat{\lambda}_1^* = 0.0219$, mentre quello per la variabile Y al valore $\hat{\lambda}_2^* = 0.0761$ (i valori esatti valgono rispettivamente $\lambda_1^* = 0.02209$ e $\lambda_2^* = 0.07476$). Il corrispondente test combinato, in modo non parametrico rispetto alla dipendenza intrinseca tra le due variabili, con la combinazione di Fisher ha dato luogo a $\hat{\lambda}_F^* = 0.0084$. Quindi l'ipotesi nulla globale viene rigettata in favore dell'alternativa di un maggiore effetto sull'ansietà e/o la depressione dovuto al maggior livello di esposizione per i lavoratori del primo reparto.

E' da notare che adottando le funzioni di combinazione di Liptak e di Tippett, i valori- p combinati sarebbero stati rispettivamente $\hat{\lambda}_L'' = 0.0072$ e $\hat{\lambda}_T'' = 0.0214$. Valori che avrebbero comunque condotto alla medesima conclusione sull'ipotesi globale. Per un commento sulla cosiddetta *combinazione iterata*, volta a rendere la conclusione finale *neutra* rispetto alla scelta più o meno arbitraria della funzione di combinazione, vedasi Pesarin (2001, pag. 153). Nel caso specifico, la combinazione iterata una sola volta dei tre test combinati porta a $\hat{\lambda}_F''' = \hat{\lambda}_L''' = \hat{\lambda}_T''' = 0.0087$. Altresì da notare è il fatto che tutti i test combinati danno luogo, in questo caso, a valori- p inferiori a quelli dei test parziali. Con ciò denotando che la combinazione non parametrica tiene conto della dipendenza tra i test parziali e dell'aggiunta di informazione che ogni test parziale offre agli altri.

3. ALCUNI PROBLEMI DI TEST MULTIPLI PER VARIABILI CATEGORIALI

In questa sezione presentiamo uno strumento utile alla individuazione della o delle classi maggiormente responsabili della significatività nella verifica dell'uguaglianza in distribuzione di una variabile categoriale nominale esaminata con riferimento a due popolazioni. Il caso di $K > 2$ popolazioni sarà ottenuto per estensione.

La soluzione classicamente adottata per la verifica d'ipotesi di uguaglianza in distribuzione è basata sul test χ^2 .

Nonostante il test χ^2 sia generalmente valido per saggiare l'ipotesi globale $H_0 : \{X_1 \stackrel{d}{=} \dots \stackrel{d}{=} X_k\}$ contro $H_1 : \{H_0 \text{ non vera}\}$, è noto che soffre di scarsa potenza e non buona approssimazione alla distribuzione asintotica nel caso di tabelle "sparse", che contengano cioè frequenze attese con valori bassi. È noto che questa situazione si presenta piuttosto frequentemente nelle applicazioni. Inoltre, nella sua formulazione standard il test χ^2 non permette di analizzare separatamente le sue componenti, ossia non permette l'analisi inferenziale a livello delle singole classi categoriali per stabilire se e quali siano maggiormente responsabili dell'eventuale significatività globale. A parziale rimedio di questa lacuna alcune soluzioni sono state proposte in letteratura, come ad esempio quelle discusse in Sham & Curtis (1995) essenzialmente basate su analisi tramite metodi di permutazione ma con controllo della significatività in generale approssimato.

In questa sezione, attraverso l'uso sistematico della metodologia di combinazione non parametrica, presentiamo uno strumento che si propone di rispondere a tali problematiche. La proposta che presentiamo consiste in un'analisi a più livelli: a livello della variabile nel suo complesso, con ciò dando luogo al cosiddetto test globale con controllo esatto della significatività rispetto alla pluralità di test, e a livello di ogni singola categoria (classe o modalità) che compone la variabile stessa con controllo della significatività in alcuni casi esatto ed in altri approssimato.

3.1. *Decomposizione dell'ipotesi globale in C sub-ipotesi parziali*

La caratteristica principale della nostra impostazione consiste nella decomposizione della tabella di contingenza in C sotto-tabelle 2×2 che verificano l'uguaglianza in distribuzione categoria per categoria nei due campioni. Ovvero: $\{(f_{jc}, n_j - f_{jc}), j = 1, 2\}$, $c = 1, \dots, C$, come in tabella 3 secondo la regola che confronta le frequenze di ogni classe rispetto a tutte le altre.

TABELLA 3

Sotto-tabella 2×2 relativa alla classe c

Gruppi	Classi		
	A_c	$\cup_{s \neq c} A_s$	
1	f_{1c}	$n_1 - f_{1c}$	n_1
2	f_{2c}	$n_2 - f_{2c}$	n_2
	f_{cC}	$n - f_{cC}$	n

In un secondo momento, i valori- p relativi alle rispettive sotto-tabelle vengono combinati in maniera non parametrica come mostrato sopra in Sezione 2. In tal modo l'ipotesi globale rimane decomposta nelle C sotto-ipotesi $H_{0c} : \{\theta_{1c} = \theta_{2c}\}$ contro $H_{1c} : \{\theta_{1c} \neq \theta_{2c}\}$, $c = 1, \dots, C$, dove $(\theta_{jc} = \Pr\{X_j \in A_c\}, c = 1, \dots, C), j = 1, 2$, rappresenta il vettore dei parametri della legge multinomiale sottostante alla popolazione j -esima. Per cui, potendo scrivere $H_0 : \{X_1 \stackrel{d}{=} X_2\} \equiv \bigcap_c H_{0c}$ e $H_1 : \{X_1 \stackrel{d}{\neq} X_2\} \equiv \bigcup_c H_{1c}$, per il problema di verifica d'ipotesi siamo nelle condizioni di poter applicare congiuntamente il principio di permutazione e la metodologia della combinazione non parametrica. Scriviamo il test parziale per la c -esima sotto-ipotesi nella forma

$$T_C^* = \left(\frac{f_{1c}^*}{n_1} - \frac{f_{2c}^*}{n_2} \right)^2 / [f_{\cdot c}(n - f_{\cdot c})].$$

Notiamo che tale test è basato sul confronto standardizzato, a meno di coefficienti costanti su tutte le permutazioni e per tutte le C sotto-tabelle. E' quindi permutazionalmente equivalente alla statistica χ^2 per la c -esima sotto-tabella 2×2 , $c = 1, \dots, C$. Inoltre è marginalmente non distorto e significativo per valori grandi. Siamo pertanto nelle condizioni previste per la corretta applicazione della loro combinazione non parametrica. Da un punto di vista strutturale, l'insieme dei test parziali corrisponde ad una comparazione classe per classe. Tra le possibili forme di combinazione, notiamo che quella diretta, ossia $T_D'' = \sum_{c \leq C} T_c$, corrisponde, essendone permutazionalmente equivalente, all'ordinario test χ^2 per la tabella $2 \times C$, per il quale va ricordato che la distribuzione è quella esatta di permutazione e non quella approssimata tramite la distribuzione asintotica. In particolare, se in luogo di T_c per la c -esima sotto-tabella si fosse usato l'equivalente test χ_c^2 (con un grado di libertà), il test globale sarebbe esattamente $\chi^2 = \sum_c \chi_c^2$, con $C - 1$ gradi di libertà, in quanto i C test parziali sono tra loro dipendenti.

Naturalmente, in luogo della combinazione diretta si potrebbe adottare una qualunque altra funzione di combinazione. In generale, pertanto si ha $T_\psi'' = \psi(\lambda_1, \dots, \lambda_C)$, dove λ_c è il valore- p di permutazione associato al test parziale T_c , $c = 1, \dots, C$. In particolare abbiamo $T_F'' = -\sum_i \log(\lambda_i)$ per la combinazione secondo Fisher, $T_T'' = \max_i(1 - \lambda_i)$ per quella secondo Tippett e $T_L'' = -\sum_i \Phi^{-1}(1 - \lambda_i)$ per quella secondo Liptak. Tra questi, a parte la combinazione diretta che dà luogo alla soluzione ordinaria, la preferenza di chi scrive va a T_T'' o più spesso a T_F'' , in quanto nelle pratiche applicazioni l'alternativa è vera solo per alcune classi, e piuttosto raramente per tutte. Inoltre, quando anche tutte le classi presentassero delle differenze, queste sarebbero solitamente di differente entità. Con la conseguenza che i relativi test parziali mostrerebbero differenti livelli di non centralità. Situazioni che giustificano la nostra preferenza, peraltro confermata dalle simulazioni

riportate nella tesi di dottorato di Finos (2003) da cui risulta che, nelle situazioni provate, T_F'' ha potenza spesso superiore a tutti gli altri, mentre non è mai risultato il peggiore.

In ogni caso, va osservato che qualunque sia la funzione di combinazione adottata, poiché tutti i test parziali sono esatti nel senso che fanno riferimento alla distribuzione di permutazione esatta, anche il test combinato è esatto, permettendo quindi un esatto controllo del rischio d'errore di prima specie per l'ipotesi globale.

La maggior ricchezza dell'analisi per sotto-ipotesi, rispetto all'analisi globale, consiste nel fatto che diviene così possibile applicare ai test parziali la metodologia delle comparazioni multiple (Hochberg and Tamhane, 1987; Westfall and Young, 1993; Hirotsu *et al.*, 2001; Finos *et al.*, 2003). Quale ulteriore ricchezza, notiamo anche che, qualora fosse di interesse per l'analisi e ricorressero le condizioni di applicabilità, tutte o parte delle sotto-ipotesi in cui il problema viene decomposto potrebbero essere di tipo direzionale. Ad esempio, potrebbe essere $H_0 : \{\bigcap_c \theta_{1c} = \theta_{2c}\}$ contro $H_{1c} : \{\bigcup_c \theta_{1c} < \neq > \theta_{2c}\}$, notazione in cui viene evidenziato che la c -esima sotto-alternativa potrebbe avere una delle forme $\{<, \neq, >\}$.

3.2. Controllo di molteplicità nell'inferenza sulle categorie

Come si è mostrato, il valore- p globale ottenuto dalla combinazione non parametrica dei valori- p parziali è un test corretto ed esatto. Per questo garantisce un controllo esatto delle significatività, nonostante i test parziali siano tra loro dipendenti. Se l'interesse inferenziale fosse ristretto alle ipotesi globali, l'analisi potrebbe essere arrestata alla valutazione globale del fenomeno. Ma se fosse ritenuto di interesse proseguire nell'analisi, l'approccio utilizzato permette di procedere ad un maggior dettaglio, utilizzando i test sulle singole sub-ipotesi parziali, ossia riferite alle singole classi o modalità. Questo permette di inferire per quali delle modalità della variabile le differenze tra i due campioni risultino significative con un controllo in parte esatto ed in parte approssimato delle significatività. Si noti che i test riferiti alle sub-ipotesi sono test separati per la verifica dell'ipotesi di uguaglianza distributiva per ogni singola modalità (ossia viene saggata esattamente quella sotto-ipotesi), mentre le sottostanti dipendenze tra i test vengono mantenute e gestite non parametricamente dalle procedure di permutazione e di combinazione, senza necessità alcuna di stimarle.

Prendiamo in considerazione la funzione di combinazione di Tippett e visualizziamo nella tabella 4 i risultati della procedura di stima della sua distribuzione di permutazione, secondo i passi e), f) e g) dell'algoritmo di simulazione della Sezione 2.2.2.

TABELLA 4
Combinazione secondo Tippett

$\hat{\lambda}_{(1)}$	$\lambda_{(1)1}^*$	$\lambda_{(1)r}^*$	$\lambda_{(1)B}^*$
T_{T_0}''	$T_{T_1}''^*$	$T_{T_r}''^*$	$T_{T_B}''^*$

in cui: $\hat{\lambda}_{(1)} = \min(\hat{\lambda}_1, \dots, \hat{\lambda}_C)$, $\lambda_{(1)r}^* = \min(\hat{L}_1(T_{1r}^*), \dots, \hat{L}_C(T_{Cr}^*))$, $T_{Tr}''^* = 1 - \lambda_{(1)r}^*$, $r = 0, 1, \dots, B$.

In base al passo g) risulta che il valore- p del test combinato, ossia $\hat{\lambda}_T'' = \sum_r \mathbf{I}(T_{Tr}''^* \geq T_{T0}'')/B$, coincide col valore- p del più grande dei test parziali calcolato con riferimento alla distribuzione dei più grandi valori tra i test parziali, propri di ogni permutazione. Pertanto, non appena si osservi che

$$\begin{aligned} \hat{\lambda}_T'' &= \sum_r \mathbf{I}(T_{Tr}''^* \geq T_{T0}'')/B = \sum_r \mathbf{I}(\lambda_{(1)r}^* \geq \hat{\lambda}_{(1)})/B = \\ &= \sum_r \mathbf{I}(\max_c T_{cr}^* \geq \max_c T_{c0})/B \end{aligned}$$

il test combinato secondo Tippett, oltre all'esatto valore- p per l'ipotesi globale, fornisce anche l'esatto valore- p di $\max_c T_c$. Questo comporta congiuntamente un controllo esatto della significatività sia del test globale e sia del più grande dei test parziali. Ciò è in parte dovuto al fatto che, nel problema in esame, tutti i test parziali hanno la medesima forma. Notiamo che, secondo la terminologia ad esempio di Westfall and Young (1993), $\hat{\lambda}_T''$ viene chiamato valore- p aggiustato per molteplicità di $\max_c T_c$, test di cui $\hat{\lambda}_{(1)}$ è il valore- p parziale.

Una correzione per molteplicità si rende necessaria anche per il secondo più grande dei test parziali, così come per gli altri successivi test ordinati.

A tal fine, una quantità di metodi con relative soluzioni approssimate sono disponibili nella letteratura corrente sull'argomento. Qui ci basta averne accennato.

3.3. Un'applicazione a dati reali

Mostriamo ora un'applicazione a dati reali della teoria fin qui esposta. Il caso considerato riguarda uno studio di associazione in ambito genetico. Si tratta di uno studio del tipo Caso/Controllo in cui la variabile dipendente è l'espressione aplotipica in un marcatore del cromosoma 17 (dati tratti da Sham and Curtis, 1995). Ai fini delle analisi che intendiamo esporre non è necessario introdurre particolari nozioni di Genetica. È rilevante soltanto il fatto che intendiamo effettuare una valutazione dell'associazione tra le frequenze aplotipiche e la variabile di classificazione Casi/Controlli, ossia di vedere se i Casi (portatori di una qualche malattia) evidenziano una distribuzione differente dai Controlli (non portatori di quella malattia) per qualcuna delle espressioni aplotipiche.

TABELLA 5

Frequenze aplotipiche per Casi e Controlli in un marcatore del cromosoma 17

Gruppo	Aplotipi										Totali
	A	B	C	D	E	F	G	H	I	J	
Casi	19	1	51	3	4	6	29	25	6	0	144
Controlli	35	1	29	6	5	3	29	38	5	2	153
Totali	54	2	80	9	9	8	58	63	11	2	297

L'associazione è stata valutata attraverso i test di $\chi^2_{\mathcal{A}}$ asintotico, T_D'' , equivalente al χ^2 esatto, T_T'' e T_F'' . Per i tre test basati su una strategia di permutazione si sono eseguite $B = 10\,000$ iterazioni Monte Carlo.

TABELLA 6
Test per l'ipotesi globale

Test	$\chi^2_{\mathcal{A}}$	T_D''	T_F''	T_T''
Val. p	.043	.031	.021	.009

TABELLA 7
Valori- p parziali e aggiustati per l'analisi sulle classi

A	B	C	D	E	F	G	H	I	J
.038	1.00	.001	.507	1.00	.324	.882	.121	.768	.492
.214	1.00	.009	.954	1.00	.873	.993	.449	.993	.935

in cui i valori- p riportati nella prima riga sono quelli parziali, mentre in seconda riga sono i corrispondenti valori aggiustati per molteplicità.

I risultati riportati in tabella 6 mostrano che tutti i test globali applicati rifiutano l'ipotesi di uguaglianza distributiva al livello $\alpha = .05$. E' comunque da sottolineare che il test combinato secondo Tippett, T_T'' , presenta un valore- p inferiore a tutti gli altri.

La tabella 7 permette di attribuire la suddetta significatività all'aplotipo C , il cui valore- p aggiustato, pari a .009 è l'unico ad essere significativo al livello $\alpha = .05$. Da sottolineare che esso coincide anche col valore- p del test globale combinato secondo Tippett, T_T'' .

La tabella 6 mostra anche che i quattro test globali applicati hanno comportamenti tra loro alquanto differenti. Ciò è dovuto alla differente forma della regione critica. In particolare, anche alla luce dei risultati dell'analisi di cui alla tabella 7, è da notare che la combinazione secondo Tippett sembra essere più efficace delle altre in quanto la diversità distributiva esistente appare attribuibile ad una sola classe, l'aplotipo C . Resta inoltre confermato che anche la combinazione secondo Fisher presenta un buon comportamento, comunque migliore del χ^2 asintotico.

Altresì è utile osservare che se ci si fosse fermati a considerare i valori- p dei test parziali, si sarebbe dovuto concludere che anche l'aplotipo A risultava significativo. Non così invece se si considera il suo valore- p aggiustato per molteplicità. Quest'ultimo, unitamente a tutti i restanti, è stato ricavato secondo la procedura "step-down" (Hochberg and Tamhane, 1987; Westfall and Young, 1993; Finos *et al.*, 2003).

3.4. Alcune estensioni uni- e multi-variate

1) L'estensione al caso di K popolazioni univariate è immediata. Infatti da una tabella di contingenza $K \times C$ possiamo estrarre $K \cdot C$ sotto-tabelle 2×2 del tipo

TABELLA 8

Sotto-tabella 2×2 relativa alla classe c della popolazione k

Gruppi	Classi		
	A_c	$\bigcup_{s \neq c} A_s$	
k^{mo}	f_{kc}	$n_k - f_{kc}$	n_k
Altri	$f_{\cdot c} - f_{kc}$	$n - n_k - f_{\cdot c} + f_{kc}$	$n - n_k$
	$f_{\cdot c}$	$n - f_{\cdot c}$	n

Poiché le ipotesi globali, con ovvio significato dei simboli, si possono scrivere come $H_0 : \{X_1 \stackrel{d}{=} \dots \stackrel{d}{=} X_K\} \equiv \bigcap_{kc} H_{0kc}$ e $H_1 : \{H_0 \text{ non vera}\} \equiv \bigcup_{kc} H_{1kc}$, per il problema di verifica d'ipotesi siamo nelle condizioni di poter applicare congiuntamente il principio di permutazione e la metodologia della combinazione non parametrica in cui i test parziali sono:

$$T_{kc}^{*2} = \left(\frac{f_{kc}^*}{n_k} - \frac{f_{\cdot c} - f_{kc}^*}{n - n_k} \right)^2 / [f_{\cdot c} (n - f_{\cdot c})], \quad 1 \leq k \leq K; 1 \leq c \leq C.$$

Ne segue che combinazione non parametrica ed analisi di molteplicità hanno le medesime connotazioni viste alla precedente sezione 3.2. Va da sé che, qualora si ritenesse di individuare prioritariamente il gruppo maggiormente responsabile dell'eventuale diversità si procederà combinando prima i test parziali di classe, gruppo per gruppo, e successivamente alla combinazione dei combinati. Viceversa, se si dovesse ritenere di interesse individuare prima la classe maggiormente responsabile. Notiamo comunque che ora, per il problema dell'analisi su $K > 2$ gruppi, il test risultante dalla combinazione diretta dei test parziali, $T_D'' = \sum_{kc \leq C} T_{kc}^{*2}$, non risulta permutazionalmente equivalente all'ordinario test globale χ^2 , pur essendone molto prossimo.

2) Per una estensione al caso multivariato in un problema a due campioni, assumiamo che la variabile risposta categoriale nominale sia Q -dimensionale, $\mathbf{X} = (X_1, \dots, X_Q)$, che il relativo numero di classi sia rappresentato nel vettore $\mathbf{C} = (C_1, \dots, C_Q)$ e che le n_1 e n_2 unità statistiche di rilevazione siano osservate indipendentemente l'una dall'altra rispettivamente dalle variabili di popolazione \mathbf{X}_1 e \mathbf{X}_2 . In questo caso le ipotesi da saggiare sono $H_0 : \{\mathbf{X}_1 \stackrel{d}{=} \mathbf{X}_2\} \equiv \{\bigcap_{q=1}^Q \bigcap_{c=1}^{C_q} (\theta_{1qc} = \theta_{2qc})\}$ verso $H_1 : \{\mathbf{X}_1 \neq \mathbf{X}_2\} \equiv \{\bigcup_{q=1}^Q \bigcup_{c=1}^{C_q} (\theta_{1qc} \neq \theta_{2qc})\}$ dove θ_{1qc} e θ_{2qc} sono i parametri della legge multinomiale della q -esima componente X_q , rispettivamente nella prima e nella seconda popolazione. Tenuto conto di quanto precede, i $\sum_q C_q$ test parziali sono

$$T_{qC_q}^{*2} = \left(\frac{f_{1qC_q}^*}{n_1} - \frac{f_{2qC_q}^*}{n_2} \right)^2 / [f_{\cdot qC_q} (n - f_{\cdot qC_q})], \quad c_q = 1, \dots, C_q, \quad q = 1, \dots, Q.$$

Per quel che concerne i passi successivi la procedura è ovviamente del tutto simile a quella vista.

Dipartimento di Scienze Statistiche
Università di Padova

FORTUNATO PESARIN

RIFERIMENTI BIBLIOGRAFICI

- A. AGRESTI, (2002), *Categorical Data Analysis (2nd edn)*, Wiley, New York.
- A. COHEN, J.H.B. KEMPERMAN, H.B. SAKROWITZ, (2000), *Properties of likelihood inference for order restricted models*, "Journal of Multivariate Analysis", 72, pp. 50-77.
- A. COHEN, D. MADIGAN, H.B. SAKROWITZ, (2003), *Effective directed tests for models with ordered categorical data*, "Australian and New Zealand Journal of Statistics", 45, pp. 285-300.
- N.A.C. CRESSIE, T.R.C. READ, (1988), *Goodness of Fit Statistics for Discrete Multivariate Data*, Springer-Verlag, New York.
- R.B. D'AGOSTINO, M.A. STEPHENS, (1986), *Goodness of Fit Techniques*, Marcel Dekker, New York.
- V. DARDANONI, A. FORCINA, (1998), *A unified approach to likelihood inference on stochastic orderings in a non-parametric context*, "Journal of the American Statistical Association", 93, pp. 1112-1123.
- R.L. DYKSTRA, S. KOCHAR, T. ROBERTSON, (1995), *Inference for likelihood ratio ordering in the two-sample problem*, "Journal of the American Statistical Association", 90, pp. 1039-1040.
- H. EL BARMÍ, R. DYKSTRA, (1995), *Testing for and against a set of linear inequality constraints in a multinomial setting*, "Canadian Journal of Statistics", 23, pp. 131-143.
- L. FINOS, (2003), *Metodi Non Parametrici per l'Analisi Multi-Focus e per il Controllo della Molteplicità con Applicazioni in Ambito Biomedico*, tesi di dottorato, Dipartimento di Scienze Statistiche, Università di Padova.
- L. FINOS, F. PESARIN, L. SALMASO, (2003), *Test combinati per il controllo della molteplicità mediante procedure di closed testing*, "Statistica Applicata, The Italian Journal of Applied Statistics", 15, pp. 301-329.
- J.L. FOLKS, (1984), *Combinations of independent tests*, in P.R. Krishnaiah and P.K. Sen (Eds.), *Handbook of Statistics*, North-Holland, Amsterdam, 4, pp. 113-121.
- C. HIROTSU, (1982), *Use of cumulative efficient scores for testing ordered alternatives in discrete models*, "Biometrika", 69, pp. 567-577.
- C. HIROTSU, (1986), *Cumulative chi-squared statistic as a tool for testing goodness of fit*, "Biometrika", 73, pp. 165-173.
- C. HIROTSU, (2002), *Multiplicity problems in the clinical trial and some statistical approaches*, in Y. Baba et Al. (eds.), *Recent Advances in Statistical Research and Data Analysis*, The Institute of Statistical Mathematics, Tokyo, pp. 1-26.
- C. HIROTSU, S. AOKI, T. INADA, T. KITAO, (2001), *An exact test for the association between the disease and alleles at highly polymorphic loci – with particular interest in the haplotype analysis*, "Biometrics", 57, pp. 148-157.
- Y. HOCHBERG, A.C. TAMHANE, (1987), *Multiple Comparisons Procedures*, Wiley, New York.
- E.L. LEHMANN, (1986), *Testing Statistical Hypotheses (2nd edn)*, Wiley, New York.
- M.D. PERLMAN, L. WU, (2002), *A defense of the likelihood ratio test for one-sided and order-restricted alternatives*, "Journal of Statistical Planning and Inference", 107, pp. 173-186.
- F. PESARIN, (1992), *A resampling procedure for nonparametric combination of several dependent tests*, "Journal of the Italian Statistical Society", 1, pp. 87-01.

- F. PESARIN, (1993), *Analisi di tabelle di contingenza multidimensionali tramite tecniche di ricampionamento*, "Archivio per l'Antropologia e la Etnologia", CXXIII, pp. 575-588.
- F. PESARIN, (1994), *Goodness of fit testing for ordered discrete distributions by resampling techniques*, "Metron", LII, pp. 57-71.
- F. PESARIN, (2001), *Multivariate Permutation Test With Application to Biostatistics*, Wiley, Chichester.
- A.R. SAMPSON, L.R. WHITAKER, (1989), *Estimation of multivariate distributions under stochastic ordering*, "Journal of the American Statistical Association", 84, 5 pp. 41-548.
- P.C. SHAM, D. CURTIS, (1995), *Monte Carlo tests for association between disease and alleles at highly polymorphic loci*, "Annals of Human Genetic", 59(1), 9, pp. 7-105.
- Y. WANG, (1996), *A likelihood ratio test against stochastic ordering in several populations*, "Journal of the American Statistical Association", 91, pp. 1676-1683.
- P.H. WESTFALL. S.S. YOUNG, (1993), *Resampling-based Multiple Testing*, Wiley, New York.

RIASSUNTO

Alcuni problemi di verifica d'ipotesi per variabili categoriali

Il lavoro considera alcuni problemi di verifica d'ipotesi per variabili categoriali multivariate in un contesto condizionale, o di permutazione. L'idea-chiave è la scomposizione delle ipotesi nulla e alternativa in un certo numero di sotto-ipotesi. Per ciascuna sotto-ipotesi si assume di disporre di un test parziale di permutazione. Naturalmente, questi test parziali non sono indipendenti. La soluzione globale si ottiene da una combinazione non parametrica dei p -value risultanti. Vengono sviluppati la teoria della combinazione non parametrica e un algoritmo computazionale per effettuare i calcoli. Come risultato particolare, è stato ottenuto un test analogo al T^2 di Hotelling per variabili nominali e ordinali. Inoltre, è delineata una analisi dal punto di vista dei test multipli al fine di stabilire quale variabile o classe o gruppo è maggiormente responsabile della significatività globale.

SUMMARY

Some hypothesis testing problems for categorical variables

This paper considers some testing problems for multivariate categorical variables within a conditional, or permutation, framework. The key idea is based on the decomposition of null and alternative hypotheses into a number of sub-hypotheses. For each sub-hypothesis it is assumed that a proper permutation partial test is available. Of course, these partial tests are assumed to be not independent. Then, the global testing solution is obtained by a nonparametric combination of resulting p -values. The theory of nonparametric combination and a computing algorithm to perform related calculations are outlined. As a particular result, a test analogous to Hotelling's T^2 for nominal and ordered categorical variables is obtained. Moreover, an analysis from the point of view of multiple testing is also outlined in order to inspect which variable or class or group is mostly responsible of a global significance.