Ricostruzione dei dati mancanti sulle unità e sul carattere in una tabella a doppia entrata

Agostino Tarsitano
Università degli studi della Calabria
Dipartimento di Economia e Statistica
87030 Arcavacata di Rende (Cs)
agotar@unical.it

Riassunto
Il lavoro propone un metodo per ricostruire i valori omessi per motivi di privacy in una tabella a doppia entrata le cui righe siano relative a delle entità di indagine e le colonne a classi di ammontare. Il metodo prevede sia l'individuazione delle unità mancanti per mezzo di un algoritmo di calcolo combinatorico che la stima dell'ammontare di carattere con il metodo di pareggiamento riga/colonna di Deming e Stephan. La percentuale di errore riscontrata nelle prove è risultata molto bassa con piena utilizzabilità degli algoritmi individuati.

keywords: valori mancanti, calcolo combonatorico, metodo Deming-Stephan

Anno 1995

Per la presente pubblicazione sono stati adempiuti gli obblighi previsti dalle norme per la consegna obbligatoria di esemplari e degli stampati delle pubblicazioni di cui alla legge del 2 febbraio 1939 n.374 e successive modificazioni.

1. Introduzione.

Il problema considerato è di ricostruire i valori mancanti in una tabella a doppia entrata relativa ad entità e classi di modalità del tipo:

Entità	$L_1 - U_1$	$L_2 - U_2$	$\dots$	$L_k - U_k$	
1	a_{11}	a_{12}		a_{1k}	$a_{1.}$
2	a_{21}	a_{22}		a_{2k}	$a_{2.}$
:					:
n	a_{n1}	a_{n2}		a_{nk}	$a_{n.}$
	$a_{.1}$	$a_{.2}$	$\dots$	$a_{.k}$	N

Entità	$L_1 - U_1$	$L_2 - U_2$	$\dots$	$L_k - U_k$	
1	s_{11}	s_{12}		s_{1k}	$s_{1.}$
2	s_{21}	s_{22}		s_{2k}	$s_{2.}$
:					:
n	s_{n1}	s_{n2}		s_{nk}	$s_{n.}$
	$s_{.1}$	$s_{.2}$	$\dots$	$s_{.k}$	T

dove a_{ij} è il numero di unità di pertinenza dell'entità i-esima e cui -singolarmente- può essere riferito un ammontare di carattere compreso nell'intervallo $[L_i$-$U_i)$. Si pensi ad esempio ad aziende industriali classificate per addetti o per classi di fatturato in varie zone produttive, ad aziende agrarie classificate per superficie agraria utilizzata a livello di comune, a contribuenti per importo versato a livello di circoscrizione comunale.

In molte occasioni capita che le entità di rilevazione costituiscano una partizione molto raffinata dei soggetti di interesse e/o che le classi di ammontare siano particolarmente dettagliate. Una possibile conseguenza è che la combinazione entità/classe potrebbe includere una sola unità: $a_{ij}=1$. In questo caso è norma che gli istituti di rilevazione non pubblichino il dato singolo troppo esposto e riconoscibile. Se questa è una prassi standard che garantisce una certa riservatezza ai vari soggetti d'indagine, non è molto confortante per i ricercatori che si trovano ad operare con dei dati incompleti (e talvolta molto incompleti). Si avverte una sensazione di disagio: ci sono dubbi ed incertezze aggiuntive sulla validità di procedure e conclusioni tratte da basi di cui è già nota la precarietà.

Per eliminare o almeno ridurre questa sgradevole carenza occorre sottoporre i dati ad un programma di correzioni logiche. In questo lavoro si propone un metodo per azzerare i differenziali di unità basato sia su degli ovvi pareggiamenti algebrici che su di un algoritmo di analisi combinatorica. Per azzerare i differenziali di ammontare si propone in una nuova applicazione un vecchio e sempre valido schema di riaggiustamento: il metodo di Deming e Stephan.

2. Simbologia e presentazione del problema

Indichiamo con a_{ij} il numero di unità pubblicato per l'entità i-esima e ricadenti nella j-esima classe e sia inoltre e a^0_{ij} il dato corretto. E' evidente che $a_{ij}=a^0_{ij}$, se $a^0_{ij}>0$ poiché quest'ultima condizione evita l'oscuramento dell'unità; se però si ha $a^0_{ij}=1$ allora, come si è detto nella introduzione, l'Istituto di Rilevazione pone $a_{ij}=0$ al fine di tutelare la riservatezza dei dati, ma negandone la completezza. L'alterazione non si estende né ai valori aggregati per entità né all'aggregazione per classi: sono quindi noti i differenziali tra dati pubblicati e dati originari ad entrambi livelli:

$$d_i = \sum_{j=1}^{k}\left(a^0_{ij} - a_{ij}\right) \qquad i = 1,2,\ldots,n$$
$$\partial_j = \sum_{i=1}^{n}\left(a^0_{ij} - a_{ij}\right) \qquad j = 1,2,\ldots,k$$

L'esclusione di un certo numero di unità implica che

$$d_i > 0 \quad \text{per} \quad i \in I \subseteq N \qquad \text{e} \qquad \partial_j > 0 \quad \text{per} \quad j \in J \subseteq K$$

per cui il confronto del numero di unità incluso nelle varie classi -tra due o più entità- ovvero, nella stessa entità, il cofronto tra classi diverse, potrebbe risultare falsato. Peraltro, e talvolta è la conseguenza più seria, si verificano delle minori attribuzioni di carattere che possono rendere molto porblematici questi confronti.

Indichiamo con s_{ij} l'ammontare di carattere attribuito alle unità riferite all'entità *i* che singolarmente ne posseggano una quantità compresa nella classe *j* e sia s^0_{ij} il corrispondente valore originario. Il fatto che $a^0_{ij}=1$ implichi che $a_{ij}=0$ comporterà la pubblicazione di $s_{ij}=0$ invece di s^0_{ij} Anche per l'ammontare di carattere sono noti i differenziali a livello di entità e di classe:

$$D_i = \sum_{j=1}^{k}\left(s^0_{ij} - s_{ij}\right) \qquad i = 1,2,\ldots,n$$
$$\Delta_j = \sum_{i=1}^{n}\left(s^0_{ij} - s_{ij}\right) \qquad j = 1,2,\ldots,k$$

La conseguenza di azzerare l'ammontare posseduto dalle unità che compaiono isolate nella combinazione entità/classe implica:

$$D_i > 0 \quad \text{per} \quad i \in I \subseteq N \qquad \text{e} \qquad \Delta_j > 0 \quad \text{per} \quad j \in J \subseteq K$$

Al fine di utilizzare dati più completi cercheremo di ricostruire i dati mancanti. Definiamo $\tilde{a}_{ij}$ i valori ricostruiti e poniamo senzaltro:

$$\left(\tilde{a}_{ij} = a_{ij}\right) \cup \left(\tilde{s}_{ij} = s_{ij}\right) \qquad \text{se} \quad a_{ij} > 1$$

Il problema consiste ora nella determinazione di quelle $\tilde{a}_{ij}$, corrispondenti a delle a_{ij}=0, che assicurino il rispetto delle seguenti condizioni:

$$\text{a)}\ \sum_{j=1}^{k} \tilde{a}_{ij} = \sum_{j=1}^{k} a_{ij}^{0} \quad i = 1, 2, \ldots, n \qquad \text{b)}\ \sum_{i=1}^{n} \tilde{a}_{ij} = \sum_{i=1}^{n} a_{ij}^{0} \quad j = 1, 2, \ldots, k$$

$$\text{c)}\ \sum_{j=1}^{k} \tilde{s}_{ij} = \sum_{j=1}^{k} s_{ij}^{0} \quad i = 1, 2, \ldots, n \qquad \text{d)}\ \sum_{i=1}^{n} \tilde{s}_{ij} = \sum_{i=1}^{n} s_{ij}^{0} \quad j = 1, 2, \ldots, k$$

$$\text{e)}\ \tilde{a}_{ij} L_j \le \tilde{s}_{ij} < \tilde{a}_{ij} U_j$$

3. Soluzione dei casi certi

Nello spirito di W.E. Deming (1940): "... The object ot taking data is to provide a basis for action, and an *adjusted value is a derived number that can be used for the purpose intended,* if it is possible to be had from the data presented for adjustment" tentiamo di aggirare il meccanismo di copertura adottato dall'Istituto di Rilevazione.

Ipotizziamo innanzitutto che i dati pubblicati siano immuni da errori e cioé escludiamo il verificarsi di condizioni del tipo:

$$\text{a)}\ (d_i = 0) \cap (D_i > 0) \qquad \text{b)}\ \left(\partial_j = 0\right) \cap \left(\Delta_j = 0\right) \qquad \text{c)}\ \partial_j < \left\lfloor \frac{\Delta_j}{U_j} \right\rfloor + 1 \ \ \text{se} \ \ \Delta_j > 0$$

Le prime due sono evidenti situazioni di errore; la condizione c) implica che, nell'ipotesi di oscuramento di qualche unità nella classe *j*, il differenziale di unità in tale classe deve prevederne almeno tante quante ne necessitano per saturare il differenziale di carattere assegnando ad ognuna il limite superiore della classe. Escludiamo inoltre dalla trattazione le entità per cui risulti $(d_i = 0) \cap (D_i = 0)$ (che non creano problemi) e concentriamoci su quelle tali che $(d_i > 0) \cup (D_i \ge 0)$.

Se $D_i = 0$ non c'è differenza tra gli ammontari pubblicati nelle singole classi e quelli originari e se i dati sono corretti una sola conclusione è possibile: $d_i = 1,\ \hat{a}_{i1} = a^0_{i1} = 1$ cioé: per l'entità i-esima è stata omessa una sola unità di pertinenza della classe nulla.

Se $D_i > 0$ occorre innanzitutto stabilire in quali classi possono ricadere le d_i unità mancanti. E' facile vedere che tali, diciamo r_i classi potenziali, sono quelle per cui è riportato un ammontare nullo di carattere $s_{ij} = 0$ e con limite inferiore più piccolo del differenziale: $L_j \le D_i$. Il verificarsi di $r_i = 1$ impone una soluzione univoca; infatti, a meno di errori, essa è compatibile solo con $d_i = 1$ e la classe di pertinenza dell'unità sarà la sola per cui si abbia: $L_h \le D_i < U_h$; si procederà pertanto all'assegnazione:

$$\hat{a}_{ih} = a^0_{ih} = 1, \quad \hat{s}_{ih} = s^0_{ih} = D_i$$

Una situazione analoga si ha se $(r_i = 1) \cap (d_i = 2)$ in quanto ora una delle unità deve rientrare nella classe nulla e l'altra dovrà essere assegnata come in precedenza:

$$\hat{a}_{i1} = a^0_{i1} = 1,\ \hat{a}_{ih} = a^0_{ih} = 1, \quad \hat{s}_{ih} = s^0_{ih} = D_i$$

Non esistono altre possibilità di allocazione simultanea certa per le unità e per il carattere a livello comunale. Se, tuttavia, ci si limita alla assegnazione delle unità rinviando ad una fase successiva quella del carattere ci sono due interessanti oportunità:

$$1)\ (r_i = d_i) \cap (a_{i1} > 0); \qquad 2)\ (r_i = d_i - 1) \cap (a_{i1} = 0)$$

In entrambi i casi è possibile eliminare il differenziale di unità senza incertezze in quanto c'è solo una combinazione di classi che può ricevere le unità mancanti.

Se non proprio delle soluzioni complete delle unità sono possibili delle assegnazioni parziali. In effetti, può succedere che una o più classi debbano forzatamente includere una delle unità mancanti. In particolare quando risulta

$$\sum_{\substack{h \in x_i \\ h \ne j}} U_h < D_i \qquad \hat{a}_{ij} = a^0_{ij} = 1$$

dove $(j_1, j_2, \ldots, j_{d_i}) = x_i$ è una combinazione delle classi potenziali dell'entità i-esima. Tale condizione potrebbe anche valere per più di una classe; addirittura, se il numero delle classi obbligate arriva a d_i si ottiene una soluzione certa per le unità.

Per le entità in cui risulta $r_i > d_i - a_{i1}$ occorrerà scegliere una delle combinazioni senza riposizione

$$\binom{r_i}{d_i} = \frac{r_i!}{d_i!(r_i - d_i)!}$$

combinazioni di classi potenziali con assegnazioni del tipo $a_{ij_1} = 1,\ a_{ij_2} = 1,\ \ldots,\ a_{ij_{d_i}} = 1$. Non tutte le x_i sono ammissibili; per esserlo debbono rispettare la condizione:

$$\sum_{h \in x_i} L_h \leq D_i < \sum_{h \in x_i} U_h$$

deve cioé essere possibile assegnare a ciascuna classe della combinazione x_i almeno il suo limite inferiore ed il totale dei limiti superiori deve essere maggiore del differenziale di carattere consentendone perciò la totale attribuzione. Le combinazioni non ammissibili sono senzaltro escluse dalla nostra trattazione e, a questo fine, definiamo

$$X_i = \left\{ x_i \middle| \sum_{h \in x_i} L_h \leq D_i < \sum_{h \in x_i} U_h \right\}$$

L'insieme di quelle combinazioni di classi potenziali che assicurano, per l'entità i-esima, sia la eliminazione del differenziale di unità che una attribuzione ragionevole del differenziale di carattere. La soluzione completa del problema delle unità sarà basato sugli insiemi X_i.

4. Calcolo di una soluzione ammissibile

Le riassegnazioni logiche dei paragrafi precedenti non assicurano una soluzione ammissibile per ciascuna entità e rispettosa dei vincoli di classe su unità e carattere. Per ottenerla è opportuno distinguere il problema delle attribuzioni di unità da quello del carattere in quanto trattabili con approcci differenti e sequenziali: il primo con uno schema combinatorico, il secondo con un algoritmo di riaggiustamento iterativo riga/colonna.

Soluzione ammissibile per le unità

Supponiamo di essere usciti dalle riassegnazioni certe con *m* comuni tali che $d_i > 0$ per $d_i > 0 \quad i = i_1, i_2, \ldots, i_m$ e sia X_i l'insieme delle soluzioni ammissibili per l'entità i-esima. La scelta di un elemento $x_i \in X_i$ garantisce l'azzeramento del differenziale di unità a livello di entità e riduce di uno lo scarto di unità a livello di classe per quelle da essa previste. Quantifichiamo l'effetto della scelta di x_i introduciamo l'indice:

$$t(x_i) = \sum_{h \in x_i} (d_h - 1)$$

che è dato dalla somma dei differenziali delle classi incluse in x_i -ridotta di d_i - ed é tanto maggiore quanto maggiori sono tali differenziali. Il problema di trovare una soluzione ammissibile per le unità può essere così formulato: determinare

$$\left(x_{i_1}, x_{i_2}, \ldots, x_{i_m}\right) \in X_{i_1} \otimes X_{i_2} \otimes \ldots \otimes X_{i_m} \ \textit{tale che} \sum_{i = i_1, i_2, \ldots, i_m} t(x_i) = 0$$

La soluzione può essere ottenuta con una ricerca esaustiva su tutti gli elementi del prodotto cartesiano $X_{i_1} \otimes X_{i_2} \otimes \ldots \otimes X_{i_m}$. I tempi di elaborazione potrebbero tuttavia risultare proibitivi. Ad esempio, per una configurazione abbastanza prudenziale quale: $d_i = 2, \ r_i = 3, \ m = 20$ sono previste 3^{20} scelte diverse. Se un computer fosse in grado di esaminarne un miliardo al secondo (velocità che oggi è strabiliante) i calcoli rischiano di essere completati non prima di un millennio.

Fortunatamente, non è necessario ricorrere alla enumerazione completa delle permutazioni: Garey e Johnson (1979) assicurano che per questo problema (da loro chiamato: *"k" largest m-tuple)* è possibile trovare un algoritmo che dia la soluzione in tempi accettabili.

Lo schema adottato in questo lavoro è un algoritmo di *backtrack* che si articola in due fasi.

Fase_1:

Siano $j_1, j_2, \ldots, j_{k'}$ le k' classi con differenziale di unità ancora positivo: $\partial_j > 0$ e siano $i_1, i_2, \ldots, i_m$ le *m* entità per cui si abbia $d_i > 0$.

a1) Per ogni $j = j_1, j_2, \ldots, j_{k'}$ si calcola ε_j il numero di entità che potenzialmente possono ricevere l'assegnazione di una ulteriore unità e cioé con

$$\left(a_{ij} = 0\right) \cap \left(L_j < D_i\right)$$

(con $\varepsilon_1 = m$) e si ordinano le classi per valori crescenti rispetto a ε_j.
b1) Sia *j* la classe corrispondente a $\varepsilon_{(1)}$. e si effettui l'assegnazione:

$$\hat{a}_{ij} = 1 \quad se \quad a_{ij} = 0 \ \ e \ \ se \sum_{\substack{j=1 \\ \hat{a}_{ij}=1}}^{k} L_j \leq D_i < \sum_{\substack{j=1 \\ \hat{a}_{ij}=1}}^{k} U_j$$

e riducendo di uno il differenziale di unità a livello di entità e di classe. Il passo b1) deve essere ripetuto per

$$i \in \left(i_1^*, i_2^*, \ldots, i_{\varepsilon_j}^*\right) \subseteq \left(i_1, i_2, \ldots, i_m\right)$$

Questo consente di procedere prima alle assegnazioni che hanno minori opportunità di allocazione.
c1) Si ripete il passo precedente per le classi corrispondenti a

$$\varepsilon_{(2)}, \varepsilon_{(3)}, \ldots, \varepsilon_{(k')}$$

La fase uno non è di solito sufficiente a risolvere in modo conveniente il problema e lascia alcune, diciamo *p*, delle *m* entità con i differenziali di unità ancora da azzerare. Queste entità sono da ritenersi più difficili a risolversi e perciò la loro sistemazione viene scelta a priori riducendo, sia pure in modo arbitrario, il problema della ricerca esaustiva.

Siano $X_i, i = i_1, i_2, \ldots, i_p$ gli insiemi di combinazioni ammissibili per entità non ancora risolte e sia x_i^* una permutazione di scelte di una data combinazione per ognuna delle *p* entità.

Fase_2:
a2) Si pone $\hat{a}_{ij} = 1$ per $i = i_1, i_2, \ldots, i_p \quad j \in x_i^*$ e si ripete l'intera fase uno dopo aver opportunamente ricalcolato *k'* ed *m*.
b2) Se non si ha ancora una soluzione completa la fase uno è ripetuta procedendo alla escussione di tutte le permutazioni in $X_{i_1} \otimes X_{i_2} \otimes \ldots \otimes X_{i_p}$
c2) Se la soluzione non è ancora raggiunta si aumenta di uno il numero di entità con soluzione vincolata e si ricomincia dal punto a2.
Nelle applicazioni pratiche il numero *p* di entità irrisolte dopo la prima fase è in genere molto ridotto e la ricerca esaustiva su tutte le loro soluzioni ha tempi praticabili. Peratro il passo c2) dovrebbe essere attivato molto di rado.

Soluzione ammissibile per le quantità di carattere

La cancellazione delle difformità tra dati pubblicati singolarmente e dati aggregati relativamente alle quantità di carattere è assimilabile al problema di stimare le probabilità congiunte di una distribuzione doppia quando siano note le probabilità marginali e sia stata rifiutata l'ipotesi di indipendenza. Tale problema ha una lunga tradizione nella letteratura statistica: Deming e Stephan (1940), Friedlander (1961), Ireland e Kullback (1968), Haberman (1979), Arthanari e Dodge (1981). Qui conviene riassumerlo adattandone opportunamente la terminologia.

Consideriamo la seguente tabella riferita a "b" entità e "c" classi riaggiustate per le unità, ma non per gli ammontari

	1	2	…	c		
1	$\hat{s}_{11}$	$\hat{s}_{12}$	…	$\hat{s}_{1c}$	$\hat{s}_{1.}$	$s^0_{1.}$
2	$\hat{s}_{21}$	$\hat{s}_{22}$	…	$\hat{s}_{2c}$	$\hat{s}_{2.}$	$s^0_{2.}$
…	…	…	…	…	…	…
b	$\hat{s}_{b1}$	$\hat{s}_{b2}$	…	$\hat{s}_{bc}$	$\hat{s}_{b.}$	$s^0_{b.}$
	$\hat{s}_{.1}$	$\hat{s}_{.2}$	…	$\hat{s}_{.c}$	$\hat{T}$	
	$s^0_{.1}$	$s^0_{.2}$	…	$s^0_{.c}$		T

dove

$$\hat{s}_{ij} = \begin{cases} L_j + \alpha & se \quad \hat{a}_{ij} = 1 \quad con \quad 0 < \alpha < \left(U_j - L_j\right) \\ 0 & se \quad \hat{a}_{ij} = 0 \end{cases}$$

e la stima iniziale delle quantità e corrisponde alle n_{ij} cioè le frquenze osservate nella tabella di contin-genza del problema a cui ci stiamo confrontando. Definiamo inoltre la quota di carattere di pertinenza della cella (i,j) come:

$$\hat{q}_{ij} = \frac{\hat{s}_{ij}}{\hat{T}} \qquad con \qquad \hat{q}_{ij} > 0; \quad \sum_{i=1}^{b}\sum_{j=1}^{c}\hat{q}_{ij} = 1$$

La stima finale della quantità di carattere q_{ij} posseduta dalle unità incluse nella cella (i,j) è il valore che minimizza la seguente funzione *(minimum discrimation function)*

$$I(q,\hat{q}) = \sum_{i=1}^{b}\sum_{j=1}^{c} q_{ij} Ln\left(\frac{q_{ij}}{\hat{q}_{ij}}\right) \quad per \;\; ogni \quad q_{ij}$$

sotto i vincoli:

$$\sum_{j=1}^{c} q_{ij} = \frac{s_{i.}^{0}}{T}; \qquad \sum_{i=1}^{b} q_{ij} = \frac{s_{.j}^{0}}{T}; \qquad q_{ij} > 0; \qquad \sum_{i=1}^{r}\sum_{j=1}^{c} q_{ij} = 1$$

Ireland e Kullback (1968) hanno dimostrato varie proprietà ottimali di questi stimatori quando il dato della cella (i,j) deriva da osservazioni empiriche più o meno soggette a fluttuazioni casuali (le n_{ij}). Nel nostro contesto esse vengono meno visto il tenore del tutto soggettivo della scelta delle s_{ij}. Non decadono però le loro proprietà matematiche e sono soprattutto queste ad interessarci. Infatti, gli stessi autori hanno ricondotto il calcolo delle q_{ij} al processo iterativo riga/colonna di Deming e Stephan semplificandone così notevolmente la determinazione. In particolare, considerando la successione:

$$\rho_j^{(1)} = 1 \qquad j = 1,2,\ldots,c \qquad\qquad \tau_i^{(1)} = \frac{q_{i.}}{\hat{q}_{i.}} \qquad i = 1,2,\ldots,b$$

$$\rho_j^{(2)} = \frac{q_{.j}}{\sum_{i=1}^{b} \tau_i^{(1)} \hat{q}_{ij}} \qquad j = 1,2,\ldots,c \qquad\qquad \tau_i^{(2)} = \frac{q_{i.}}{\sum_{j=1}^{c} \rho_j^{(2)} \hat{q}_{ij}} \qquad i = 1,2,\ldots,b$$

…..........................

$$\rho_j^{(n)} = \frac{q_{.j}}{\sum_{i=1}^{b} \tau_i^{(n-1)} \hat{q}_{ij}} \qquad j = 1,2,\ldots,c \qquad\qquad \tau_i^{(n)} = \frac{q_{i.}}{\sum_{j=1}^{c} \rho_j^{(n)} \hat{q}_{ij}} \qquad i = 1,2,\ldots,b$$

Ireland e Kullback dimostrano che:

$$\lim_{n\to\infty} q_{ij}^{(n)} = \lim_{n\to\infty} \tau_i^{(n)} \rho_j^{(n+1)} \hat{q}_{ij} = q_{ij}$$

con una convegenza molto rapida. Resta da stabilire se le stime finali ottenute siano comprese tra il limite inferiore e superiore delle classi e cioé

$$\frac{L_j}{T} \leq q_{ij}^{(n)} < \frac{U_j}{T}$$

La teoria del procedimento di Deming e Stephan non prevede delle condizioni che portino a stime sicuramente incluse nei limiti. Poiché tali sforamenti non sono infrequenti è necessario ovviarvi ed in questo senso si è scelto di porre l'ammontare di carattere della cella con valori non ammissibili, pari al limite violato:

$$\tilde{s}_{ij} = \begin{cases} L_j & \text{se} \quad \tilde{s}_{ij}^{(n)} < L_j \\ U_j & \text{se} \quad \tilde{s}_{ij}^{(n)} > U_j \\ \tilde{s}_{ij}^{(n)} & \text{se} \quad L_j \le \tilde{s}_{ij}^{(n)} \le U_j \end{cases}$$

e ricominciando da capo il riaggiustamento con questi nuovi valori iniziali. In effetti, tale scelta riduce i differenziali di caratteri da ripartire, sia a livello di riga che di colonna, e -dopo un certo numero di iterazioni- si ha quasi sempre la convergenza a valori ammissibili (almeno entro un certo grado di approssimazione, diciamo dello 0.01)

6. Esempio applicativo

I dati sulle aziende agrarie (superficie agraria utilizzata o SAU e numero di aziende) di un generico comune sono organizzate dall'ISTAT nelle seguente classificazione

Tabella_1: classi* di SAU

	Aziende		SAU	
	L_i	U_i	L_i	U_i
1)	0	0	a_{i1}	0
2)	0†	1	a_{i2}	s_{i2}
3)	1	2	a_{i3}	s_{i3}
4)	2	5	a_{i4}	s_{i4}
5)	5	10	a_{i5}	s_{i5}
6)	10	20	a_{i6}	s_{i6}
7)	20	50	a_{i7}	s_{i7}
8)	50	100	a_{i8}	s_{i8}
9)	100	∞**	a_{i9}	s_{i9}
Tot			$a_{i.}$	$s_{i.}$

* gli estremi inferiori si intendono inclusi e quelli superiori esclusi.
† In realtà si è considerato un limite pari a L_1=0.01;
** Nelle elaborazioni si è posto $U_8=10^9$;

Per ogni classe l'istituto di rilevazione fornisce, sia a livello di comune che a livello di classe, il numero di unità e, tranne che per la prima classe (senza terreno agrario), la SAU complessiva delle aziende ricadenti nella classe. A tutela del segreto statistico sono però eliminati i dati relativi a singole aziende. Nella prima pagina in appendice sono riportate due tabelle con i dati sulle aziende agrarie e sulla SAU limitatamente ai comuni interessati da oscuramenti non risolvibili con le procedure certe del paragrafo 3.

Tabella_1.a: dati pubblicati con oscuramenti per le aziende agrarie in provincia di Catanzaro . Censimento '90

	SS	0-1	1-2	2-5	5-10	10-20	20-50	50-100	>100	$a_{i.}$	$a^0_{i.}$	d_i
1	6	252	95	67	12	5	0	0	0	437	439	2
2	0	250	82	33	5	4	0	0	0	374	376	2
3	0	258	102	51	16	7	0	0	0	434	437	3
4	0	214	62	30	0	4	0	0	0	310	312	2
5	0	89	69	64	17	6	0	9	0	254	257	3
6	0	255	98	77	6	0	0	0	0	436	438	2
7	0	359	184	81	4	3	0	0	0	631	633	2
8	0	191	97	57	10	4	0	0	0	359	362	3
9	0	64	67	51	26	3	0	0	0	211	213	2
10	0	161	75	51	13	0	0	0	0	300	302	2
11	0	258	67	33	0	0	3	0	0	361	363	2
12	0	108	36	13	0	2	0	0	0	159	162	3
13	0	150	54	36	10	0	0	0	0	250	253	3
14	5	247	91	52	15	8	6	0	0	424	426	2
15	0	355	150	95	13	0	0	0	0	613	615	2
16	0	163	128	160	84	17	6	0	0	558	561	3
17	0	218	140	125	35	4	8	0	0	530	532	2
18	0	91	46	53	23	5	0	0	0	218	221	3
19	2	347	173	73	17	9	0	0	0	621	624	3
20	0	93	54	49	14	4	0	0	0	214	216	2
21	3	337	31	12	0	0	2	0	0	385	388	3
22	0	37	55	150	73	22	13	0	0	350	353	3
23	0	160	87	65	11	10	6	0	0	339	342	3
24	2	44	42	76	18	2	3	0	0	187	189	2
25	0	86	35	28	4	0	0	0	0	153	156	3
26	0	167	57	43	17	8	7	0	0	299	301	2

Tabella_1.b: dati pubblicati con oscuramenti per le SAU in provincia di Catanzaro- '90

	0-1	1-2	2-5	5-10	10-20	20-50	50-100	>100	$s_{i.}$	$s^0_{i.}$	D_i
1	124.78	125.73	208.98	77.93	70.20	0.00	0.00	0	607.62	735.49	127.87
2	127.45	115.50	97.60	34.08	44.10	0.00	0.00	0	418.73	509.47	90.74
3	106.66	131.40	135.94	108.00	99.64	0.00	0.00	0	581.64	733.64	152.00
4	93.17	83.21	81.01	0.00	53.34	0.00	0.00	0	310.73	437.95	127.22
5	42.2	81.82	198.24	113.92	73.13	0.00	693.85	0	1203.16	1525.16	322.00
6	123.76	144.64	221.16	35.10	0.00	0.00	0.00	0	524.66	1084.86	560.20
7	177.24	252.00	211.42	26.60	38.20	0.00	0.00	0	705.46	883.07	177.61
8	85.73	137.80	157.51	64.40	66.23	0.00	0.00	0	511.67	797.05	285.38
9	37.43	95.24	151.62	185.62	46.22	0.00	0.00	0	516.13	624.13	108.00
10	69.18	104.02	150.24	88.59	0.00	0.00	0.00	0	412.03	452.60	40.57
11	102.51	87.80	93.64	0.00	0.00	109.00	0.00	0	392.95	412.45	19.50
12	42.17	51.48	38.38	0.00	26.00	0.00	0.00	0	158.03	351.03	193.00
13	76.93	71.92	103.33	68.88	0.00	0.00	0.00	0	321.06	449.95	128.89
14	109.47	121.08	140.68	109.66	113.61	237.50	0.00	0	832.00	1002.00	170.00
15	170.82	208.26	285.36	84.08	0.00	0.00	0.00	0	748.52	786.64	38.12
16	97.01	181.35	514.05	569.39	210.85	192.35	0.00	0	1765.00	2078.00	313.00
17	123.79	208.96	404.57	229.02	47.88	230.87	0.00	0	1245.09	1421.59	176.50
18	43.3	56.33	164.91	154.05	58.38	0.00	0.00	0	476.97	616.59	139.62
19	159.87	235.81	197.00	106.64	112.41	0.00	0.00	0	811.73	987.93	176.20
20	46.22	71.29	145.00	90.53	55.01	0.00	0.00	0	408.05	507.05	99.00
21	123.74	40.73	36.20	0.00	0.00	81.19	0.00	0	281.86	482.85	200.99
22	19.17	78.18	473.69	496.56	288.38	355.63	0.00	0	1711.61	2132.27	420.66
23	74	114.22	201.44	83.30	133.13	191.70	0.00	0	797.79	1016.22	218.43
24	26.58	58.08	245.15	125.65	31.20	98.00	0.00	0	584.66	934.39	349.73
25	38.71	42.06	74.35	22.87	0.00	0.00	0.00	0	177.99	215.58	37.59
26	63.9	74.09	127.27	103.93	101.89	208.30	0.00	0	679.38	1009.46	330.08

Tabella_2.a: dati stimati per le aziende agrarie in provincia di Catanzaro

	SS	0-1	1-2	2-5	5-10	10-20	20-50	50-100	>100	$a_{i.}^{0}$
1	6	252	95	67	12	5	1	0	1	439
2	0	250	82	33	5	4	1	1	0	376
3	1	258	102	51	16	7	1	0	1	437
4	0	214	62	30	0	4	1	0	1	312
5	1	89	69	64	17	6	1	9	1	257
6	0	255	98	77	6	1	0	0	1	438
7	0	359	184	81	4	3	1	0	1	633
8	0	191	97	57	10	4	1	1	1	362
9	0	64	67	51	26	3	1	1	0	213
10	0	161	75	51	13	1	1	0	0	302
11	0	258	67	33	1	1	3	0	0	363
12	0	108	36	13	1	2	1	0	1	162
13	0	150	54	36	10	1	1	1	0	253
14	5	247	91	52	15	8	6	1	1	426
15	0	355	150	95	13	1	1	0	0	615
16	1	163	128	160	84	17	6	1	1	561
17	0	218	140	125	35	4	8	1	1	532
18	1	91	46	53	23	5	1	1	0	221
19	2	347	173	73	17	9	1	1	1	624
20	1	93	54	49	14	4	0	1	0	216
21	3	337	31	12	1	0	2	1	1	388
22	1	37	55	150	73	22	13	1	1	353
23	1	160	87	65	11	10	6	1	1	342
24	2	44	42	76	18	2	3	1	1	189
25	1	86	35	28	4	1	1	0	0	156
26	0	167	57	43	17	8	7	1	1	301

Tabella_2.b: dati stimati per le SAU in provincia di Catanzaro

	0-1	1-2	2-5	5-10	10-20	20-50	50-100	>100	$s_{i.}^{0}$
1	124.78	125.73	208.98	77.93	70.20	20.12	0.00	107.75	735.49
2	127.45	115.50	97.60	34.08	44.10	32.62	58.13	0.00	509.47
3	106.66	131.40	135.94	108.00	99.64	21.72	0.00	130.28	733.64
4	93.17	83.21	81.01	0.00	53.34	20.10	0.00	107.12	437.95
5	42.2	81.82	198.24	113.92	73.13	46.01	693.85	275.99	1525.16
6	123.76	144.64	221.16	35.10	19.87	0.00	0.00	540.33	1084.86
7	177.24	252.00	211.42	26.60	38.20	25.38	0.00	152.23	883.07
8	85.73	137.80	157.51	64.40	66.23	32.50	57.92	194.95	797.05
9	37.43	95.24	151.62	185.62	46.22	38.82	69.18	0.00	624.13
10	69.18	104.02	150.24	88.59	14.32	26.25	0.00	0.00	452.60
11	102.51	87.80	93.64	6.87	12.64	109.00	0.00	0.00	412.45
12	42.17	51.48	38.38	7.84	26.00	26.46	0.00	158.70	351.03
13	76.93	71.92	103.33	68.88	19.96	39.16	69.78	0.00	449.95
14	109.47	121.08	140.68	109.66	113.61	237.50	50.10	119.90	1002.00
15	170.82	208.26	285.36	84.08	13.46	24.67	0.00	0.00	786.64
16	97.01	181.35	514.05	569.39	210.85	192.35	71.70	241.30	2078.00
17	123.79	208.96	404.57	229.02	47.88	230.87	50.12	126.38	1421.59
18	43.3	56.33	164.91	154.05	58.38	49.78	89.84	0.00	616.59
19	159.87	235.81	197.00	106.64	112.41	20.12	50.01	106.07	987.93
20	46.22	71.29	145.00	90.53	55.01	0.00	99.00	0.00	507.05
21	123.74	40.73	36.20	7.10	0.00	81.19	50.16	143.73	482.85
22	19.17	78.18	473.69	496.56	288.38	355.63	95.05	325.61	2132.27
23	74	114.22	201.44	83.30	133.13	191.70	50.03	168.40	1016.22
24	26.58	58.08	245.15	125.65	31.20	98.00	80.11	269.62	934.39
25	38.71	42.06	74.35	22.87	13.27	24.32	0.00	0.00	215.58
26	63.9	74.09	127.27	103.93	101.89	208.30	75.61	254.47	1009.46