

# I fondamenti dell'analisi statistica implicativa

Regis Gras<sup>1</sup>

## Riassunto

Parte delle situazioni di didattica dove si trattava di gerarchizzare problemi in funzione delle difficoltà avvertite dagli allievi, il metodo implicativo si sviluppa sulla scia dei problemi che essa incontra, delle questioni che vengono poste. Problemi sui quali pensa poter agire al fine di strutturare e permettere, a partire dalla contingenza delle regole, di esplicitare e dunque di prevedere nei differenti domini: la psicologia, la sociologia, la biologia, tra le altre. E' così che i concetti d'intensità d'implicazione, di coesione di classi, di misura di implicazione-inclusione, di significatività di livelli gerarchici, di contributi di variabili supplementari, etc., sono stati sviluppati. E' anche per rispondere che al trattamento delle variabili binarie si sono aggiunte quelle delle variabili modali, frequenziali e, più recentemente, variabili di intervallo.

I trattamenti automatici dei calcoli e dei grafici sono eseguiti con l'aiuto del software C.H.I.C. (Classification Hiérarchique Implicative et Cohésive) disponibile sotto Windows 95<sup>2</sup>.

## Résumé

Partie de situations de didactique des mathématiques où il s'agissait de hiérarchiser des problèmes en fonction des difficultés ressenties par les élèves, la méthode implicative se développe au fil des problèmes qu'elle rencontre, des questions qu'ils posent. Problèmes sur lesquels elle pense pouvoir agir afin de structurer et permettre, à partir de la contingence de règles, d'expliquer et donc de prévoir dans différents domaines : la psychologie, la sociologie, la biologie, entre autres. C'est ainsi que les concepts d'intensité d'implication, de cohésion de classes, de mesure d'implication-inclusion, de significativité de niveaux de hiérarchie, de contribution de variables supplémentaires, etc., ont été développés. C'est aussi pour y répondre qu'au traitement de variables binaires se sont ajoutés ceux des variables modales, fréquentielles et, plus récemment, de variables-intervalles.

Les traitements automatiques des calculs et des graphiques sont exécutés à l'aide du logiciel C.H.I.C. (Classification Hiérarchique Implicative et Cohésive) disponible sous Windows 95<sup>3</sup>.

---

<sup>1</sup> Conferenza tenuta all'Ecole Polytechnique dell'Università di Nantes. Il testo in francese si trova nel sito INTERNET del GRIM: <http://dipmat.math.unipa.it/~grim>.

<sup>2</sup> La versione per Windows 95 è sviluppata da Raphael COUTURIER.

<sup>3</sup> La version pour Windows 95 est développée par Raphaël COUTURIER

## § 1 L'intensità d'implicazione statistica rivisitata.

### 1.1 Ricordiamo la problematica dell'implicazione statistica nel caso binario.

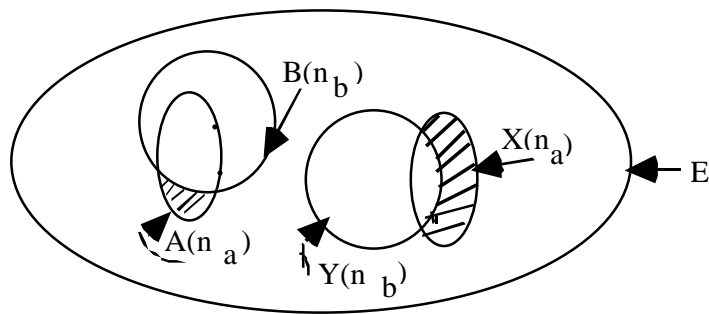
Una popolazione  $E$  di oggetti o soggetti è incrociata con delle variabili (caratteri, criteri, riuscite,...) di un insieme che lo si interroga nella maniera seguente: “in quale misura si può considerare che il fatto di rilevare la variabile  $a$  implica quello di rilevare la variabile  $b$ ? Detto altrimenti, gli oggetti hanno la tendenza ad essere  $b$  se si sa che sono  $a$ ?” Nelle situazioni naturali, umane o scienze di vita, dove i teoremi (se  $a$  allora  $b$ ) nel senso matematico del termine non possono essere stabilite dalle eccezioni che le contaminano, è importante per il ricercatore e l'utilizzatore di “scavare nei suoi dati” al fine di liberare però delle regole molto deboli per poter descrivere, strutturare una popolazione e congetturare una certa stabilità a dei fini predicativi. Ma questo scavo esige la messa a punto di metodi per guidarla e per liberarla dall'andare a tantonismi e dall'empirismo.

Per questo, allo stesso modo del metodo di misura di similarità di I.C. Lermann (LERMANN I.C. 1970, 1981), definiamo (GRAS R. 1979, 1996) la misura della relazione implicativa  $a \rightarrow b$  a partire dall'inversomiglianza dell'apparizione, nei dati, dal numero di casi che l'invalidano, per i quali  $a$  è verificato senza che  $b$  lo sia. Questa misura è relativizzata al numero dei dati verificanti rispettivamente  $a$  e non  $b$ . Essa quantifica “lo stupore” dell'esperto davanti il numero inverosimilmente piccolo di contro esempi dovuti ad una indipendenza presunta ed agli effettivi in gioco.

Precisiamo. Un insieme finito  $V$  di  $v$  variabili è dato:  $a, b, c, \dots$ . Nella situazione pragmatica classica, si tratta di “performances” (vero-falso) a degli items di un questionario. Ad un insieme finito  $E$  di  $n$  soggetti  $x$ , si associano, per abuso di scrittura, le funzioni del tipo:  $x \rightarrow a(x)$  dove  $a(x)=1$  ( $a(x)=\text{vero}$ ) se  $x$  soddisfa o possiede il carattere  $a$  e  $0$  ( $a(x)=\text{falso}$ ) in caso contrario. In intelligenza artificiale, si dirà che  $x$  è un esempio o una istanza per  $a$  se  $a(x)=1$  ed un contro esempio nel caso contrario.

La regola “ $a \rightarrow b$ ” è logicamente vera se per ogni  $x$ ,  $b(x)$  non è nullo che nel caso dove  $a(x)$  lo è anche; detto altrimenti se l'insieme  $A$  degli  $x$  per i quali  $a(x)=1$  è contenuto nell'insieme  $B$  degli  $x$  per i quali  $b(x)=1$ . Tuttavia, questa inclusione stretta non è osservata nella realtà che eccezionalmente. Nel caso di un questionario di conoscenze, si potrà in effetti osservare alcuni rari allievi che riescono nell'item  $a$  e non riescono nell'item  $b$ , senza che sia contestata la tendenza ad avere  $b$  quando si ha  $a$ . Relativamente ai cardinali di  $E$  (sia  $n$ ), ma anche di  $A$  (sia  $n_a$ ) e  $B$  (sia  $n_b$ ), è dunque il “peso” dei contro esempi (sia  $n_a - n_b$ ) che bisognerà prendere in conto per accettare statisticamente di conservare o non la **quasi-implicazione** o la **quasi regola** “ $a \rightarrow b$ ”.

Per matematizzare questo, consideriamo, come fa I.C.Lerman per la similarità, due parti qualunque  $X$  e  $Y$  di  $E$ , scelte aleatoriamente ed indipendentemente (assenza di legami priori tra queste due parti) e degli stessi cardinali rispettivi che  $A$  e  $B$ . Sia  $\bar{Y}$  e  $\bar{B}$  i complementari rispettivi di  $Y$  e di  $B$  in  $E$  di cardinale  $n_{\bar{Y}}$ .



Le parti tratteggiate  
Corrispondenti alla  
non-soddisfazione di  
implicazione di a su b

Noi diremo allora:

**Definizione 1**

$a \Rightarrow b$  è ammissibile al livello di confidenza  $1-\alpha$  se e solo se

$$\Pr[\text{card}(X \cap \bar{Y}) \leq \text{card}(A \cap \bar{B})]$$

Noi dimostriamo (LERMAN I.C. GRAS R., ROSTAM H., 1981) che, per un certo processo di trascinamento

La variabile aleatoria  $\text{Card}(X \cap \bar{Y})$  segue la legge di Poisson del parametro  $\frac{n_a n_{\bar{b}}}{n}$ .

Nel caso dove  $n_{\bar{b}} > 0$ , noi riduciamo e centriamo questa variabile di Poisson nella variabile:

$$Q(a, \bar{b}) = \frac{\text{Card}(X \cap \bar{Y}) - n p_a p(\bar{b})}{\sqrt{n p_a p(\bar{b})}} = \frac{\text{Card}(X \cap \bar{Y}) - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$$

Nella realizzazione sperimentale, il valore osservato di  $Q(a, \bar{b})$  è  $q(a, \bar{b})$

**Definizione 2**

$$q(a, \bar{b}) = \frac{n_a \bar{b} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} \text{ chiamato indice di implicazione.}$$

Numero che riteniamo come indicatore della non-implicazione di a su b.

Nei casi che legittimano convenientemente l'approssimazione (per esempio,  $\frac{n_a n_{\bar{b}}}{n} > 3$ ), la variabile  $Q(a, \bar{b})$  segue approssimativamente la legge normale centrata ridotta. L'intensità di implicazione, qualità dell'ammissibilità di a su b, per  $n_a > n_b$  e  $n_b \approx n$ , è allora definita a partire dell'indice  $q(a, \bar{b})$  da:

### Definizione 3

Nel caso dove  $n_b \neq n$ , l'intensità di implicazione di a su b è:

$$q(a, \bar{b}) = 1 - \Pr[Q(a, \bar{b}) \leq q(a, \bar{b})] = \frac{1}{\sqrt{2}} \int_{q(a, \bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$$

In seguito, la definizione di implicazione diviene:

### Definizione 4

L'implicazione  $a \Rightarrow b$  sarà ammissibile al livello di confidenza  $1-\alpha$  se e solo se

$$q(a, \bar{b}) = 1 - \Pr[Q(a, \bar{b}) \leq q(a, \bar{b})] \geq 1 - \alpha$$

Ricordiamo che questa modellizzazione della quasi-implicazione misura lo stupore di constatare la piccolezza dei contro esempi rispetto al numero sorprendente delle istanze di implicazione. Per conseguenza, se la regola è banale, come nel caso dove B è molto grande o coincide con E, questo stupore diviene piccolo. Dimostriamo che questa banalità si traduce in una intensità d'implicazione molto debole ed anche nulla:

**Se,  $n_a$  essendo fissato in A essendo incluso in B,  $n_b$  tende verso n (B "cresce" verso E), allora  $q(a, \bar{b})$  tende verso 0.**

**Nota 1:** Altre modellizzazioni, oltre quella di Poisson, sono possibili. Citiamo:

a) *una modellizzazione binomiale:* considerando le variabili duali  $\text{card}(A \cap \bar{Y})$  e  $\text{card}(X \cap \bar{B})$ , dove X e Y sono delle parti scelte in maniera indipendente in E e rispettanti le proprietà cardinali rispettive di A e B, ogni elemento di E, per esempio, ha la probabilità  $\frac{n_a}{n} \frac{n_b}{n}$  di appartenere a  $A \cap \bar{Y}$ . Come segue:

$$\Pr[\text{card}(A \cap \bar{Y}) = k] = C_n^k \left(\frac{n_a n_b}{n^2}\right)^k \left(1 - \frac{n_a n_b}{n^2}\right)^{n-k} = \Pr[\text{card}(X \cap \bar{B}) = k]$$

• una modellizzazione ipergeometrica: la si può vedere rapidamente considerando ancora le variabili aleatorie  $\text{card}(A \cap \bar{Y})$  e  $\text{card}(X \cap \bar{B})$ . Si ha in effetti:

$$\begin{aligned} \Pr[\text{card}(A \cap \bar{Y}) = k] &= \frac{C_{n_a}^k C_{n-n_a}^{n-n_b-k}}{C_n^{n-n_b}} = \frac{n_a! n_{\bar{a}}! n_b! n_{\bar{b}}!}{k!(n_a - k)!(n_{\bar{b}} - k)!(n_{\bar{b}} - n_a - k)!} \\ &= \frac{C_{n-n_b}^k C_{n_a-k}^{n_a-k}}{C_n^{n_a}} = \Pr[\text{card}(X \cap \bar{B}) = k] \end{aligned}$$

**Nota 2:** La quasi-implicazione, di indice  $q(a, \bar{b})$  non simmetrica, non coincide con il coefficiente di correlazione  $r(a, b)$  che è simmetrico e che rende conto del legame tra le variabili  $a$  e  $b$ . In effetti noi mostriamo che se  $q(a, \bar{b}) = 0$  allora  $\frac{r(a, b)}{q(a, \bar{b})} = -\sqrt{\frac{n}{n_b n_{\bar{a}}}}$

**Nota 3:** Noi possiamo definire delle congiunzioni di variabili del tipo "a e b" o "(a e b) o c..." al fine di modellizzare i fenomeni che rilevano dei concetti come viene fatto nell'apprendimento o in intelligenza artificiale. I calcoli associati restano compatibili con la logica delle proposizioni collegate da connettivi.

**Nota 4:** Contrariamente all'indice di Loewinger (1942) e alla probabilità condizionale  $(Pr[B/A])$  e tutte le sue derivate, l'intensità di implicazione varia con la dilatazione degli insiemi  $E, A$  e  $B$ , questo non può che rendere statisticamente credibile la relazione che vogliamo modellizzare.

### 1-2 Casi di variabili modali e frequenziali

Nel seguito dei nostri lavori, estendiamo la nozione di implicazione statistica a delle variabili non binarie. E' il caso delle variabili modali che sono associate a dei fenomeni dove i valori  $a(x)$  sono dei numeri dell'intervallo  $[0,1]$  e che descrivono dei gradi di appartenenza o di soddisfazione come in logica sfumata. E' anche il caso delle variabili frequenziali che sono associate a dei fenomeni dove i valori di  $a(x)$  sono dei reali positivi qualunque.

J. B. Lagrange (1998) ha dimostrato che, nel caso modale,

- se  $a(x)$  e  $\bar{b}(x)$  sono i valori presi in  $x$  dalle variabili modali  $a$  e  $\bar{b}$ , con  $\bar{b}(x) = 1 - b(x)$

- se  $s_a^2$  e  $s_{\bar{b}}^2$  sono le varianze empiriche delle variabili  $a$  e  $\bar{b}$

allora l'indice di implicazione, che egli chiama indice di propensione, diventa:

#### Definizione 5

$$q(a, \bar{b}) = \frac{\sum_{x \in E} a(x) \bar{b}(x) - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{(n^2 s_a^2 + n_a^2)(n^2 s_{\bar{b}}^2 + n_{\bar{b}}^2)}{n}}} \text{ è l'indice di propensione della variabile modale}$$

Egli prova egualmente che questo indice coincide con l'indice definito precedentemente nel caso binario se il numero della modalità di  $a$  e di  $b$  è giustamente 2, perché in questo caso:

$$n^2 s_a^2 + n_a^2 = n n_a, \quad n^2 s_{\bar{b}}^2 + n_{\bar{b}}^2 = n n_{\bar{b}} \quad \text{e} \quad \sum_{x \in E} a(x) \bar{b}(x) = \frac{n_a n_{\bar{b}}}{n}.$$

Questa soluzione riportata al caso modale è anche applicabile al caso delle variabili frequenziali, ed anche a delle variabili numeriche positive, alla condizione di avere

normalizzato i valori osservati sulle variabili, tali che a e b, la normalizzazione in [0,1] essendo fatta a partire dal massimo del valore preso rispettivamente da a e b sull'insieme E.

### 1-3 Caso delle variabili-intervallo

#### 1-3-1 Situazione fondamentale

Due variabili reali a e b prendono un certo numero di valori su due intervalli finiti  $[ \alpha_1, \alpha_2 ]$  e  $[ \beta_1, \beta_2 ]$ . Sia A (risp. B) l'insieme dei valori di a (risp b) osservati su  $[ \alpha_1, \alpha_2 ]$  (risp  $[ \beta_1, \beta_2 ]$ ). Per esempio, a rappresenta i pesi di un insieme di n soggetti e b le taglie di questi stessi soggetti.

Si pongono due problemi:

1° possiamo definire dei sotto intervalli adiacenti  $[ \alpha_1, \alpha_2 ]$  (risp  $[ \beta_1, \beta_2 ]$ ) affinché la partizione la più fine ottenuta rispetti al meglio la distribuzione osservata in  $[ \alpha_1, \alpha_2 ]$  (risp  $[ \beta_1, \beta_2 ]$ )?

2° possiamo trovare le partizioni rispettive di  $[ \alpha_1, \alpha_2 ]$  e  $[ \beta_1, \beta_2 ]$  costituite da unioni di sotto intervalli adiacenti precedenti, partizioni che massimizzano l'intensità di implicazione media dei sotto intervalli dell'uno su dei sotto intervalli sull'altro appartenente a queste partizioni?

Cercheremo di rispondere a queste due questioni scegliendo dei criteri per ottimizzare e per soddisfare l'ottimalità attesa in ogni caso.

#### 1-3-1 Primo problema

Ci si interesserà all'intervallo  $[ \alpha_1, \alpha_2 ]$  supponendolo munito di una partizione iniziale banale di sotto intervallo della stessa lunghezza, ma non necessariamente della stessa distribuzione di frequenze osservate su questi sotto intervalli.

Notiamo  $P_0 = \{A_{01}, A_{02}, \dots, A_{0p}\}$ , questa partizione in p sotto intervalli. Si cerca di ottenere una partizione  $P_q^*$  di  $[ \alpha_1, \alpha_2 ]$  in p sotto intervalli  $A_{q1}, A_{q2}, \dots, A_{qp}$  in modo tale che all'interno di ogni sotto intervallo si abbia una buona omogeneità statistica (debole inerzia intra-classe) e che questi sotto intervalli presentano una buona mutua eterogeneità (forte inerzia inter-classe). Si sa che se uno dei criteri è verificato l'altro lo è necessariamente (teorema di Koenig-Huyghens).

Per fare ciò, si adopererà un metodo direttamente ispirato dal metodo delle nuvole dinamiche concepito da Edwin Diday<sup>4</sup> ed adattato alla situazione presente. Per questo, si cerca di minimizzare una certa funzione W, definita sull'insieme G dei punti reali di  $[ \alpha_1, \alpha_2 ]$  e l'insieme delle partizioni P di  $[ \alpha_1, \alpha_2 ]$  in p sotto intervalli di  $A_i$ , nel modo seguente:

$$W(G, P) = \sum_{i=1}^p D(G_i, A_i) \text{ avec } D(G_i, A_i) = \int_{x \in A_i} (G_i - x)^2 \text{ pour tout } i=1,2,\dots,p.$$

Così se G è il baricentro dei valori osservati in A, se  $G_i$  è il baricentro dei valori osservati in  $A_i$  allora  $W(G, P)$  è l'inerzia intra-classe di A e  $D(G_i, A_i)$  è l'inerzia di  $A_i$ .

<sup>4</sup> Diday (E), *Nouvelles méthodes et nouveaux concepts en classification automatique et reconnaissance des formes*, thèse de doctorat d'état, Université de Paris VI, 1972

### 1<sup>a</sup> tappa

Si parte dalla partizione  $P_0 = \{A_{01}, A_{02}, \dots, A_{0p}\}$ . Si sceglie quello che E. Diday chiama nodo, in numero  $p$ , in  $[1, 2]$ . Questi nodi sono scelti confusi con i rispettivi baricentri, tali che  $G_{1i}$ , dei sotto intervalli  $A_{0i}$  dei valori che sono osservati. Sia  $G_1$  il loro baricentro.

Si cerca allora la partizione  $P_1 = \{A_{11}, A_{12}, \dots, A_{1p}\}$  tale che per ogni  $i$ :

$$A_{1i} = \left\{ x \mid \sum_j (G_{1i} - x)^2 (G_{1j} - x)^2 \right\}$$

Questo ci porta a costituire  $A_{1i}$  con l'aiuto dei punti che sono i più vicini al baricentro di  $A_{0i}$ . Questo porta anche a riportare nei sotto intervalli vicini i punti che sono più vicini dei loro propri baricentri rispettivi. In caso di uguaglianza, si assume  $x$  al sotto intervallo del più piccolo indice.

Si dimostra allora che  $W(G_1, P_1) \leq W(G_1, P_0)$

### 2<sup>a</sup> tappa

Si dispone della partizione:  $P_1 = \{A_{11}, A_{12}, \dots, A_{1p}\}$

Si sceglie  $p$  nodi  $G_{2i}$  che minimizzano rispettivamente le quantità  $D(y, A_{1i})$ , cioè: per ogni  $y$ , valore osservato in  $A$ ,  $D(G_{2i}, A_{1i}) = \sum_{x \in A_{1i}} (y - x)^2$ , che è l'inerzia di  $A_{1i}$  attorno ad  $y$

con un coefficiente vicino. Di conseguenza, il nodo  $G_{2i}$  è il baricentro di  $A_{1i}$ .

Si ottiene dunque una nuova successione di nodi:  $\{G_{21}, G_{22}, \dots, G_{2p}\}$  di cui  $G_2$  è il baricentro e lo si riporta al procedimento della tappa precedente.

Si determina in effetti la partizione  $P_2 = \{A_{21}, A_{22}, \dots, A_{2p}\}$  tale che:

$$A_{2i} = \left\{ x \mid \sum_j (G_{2i} - x)^2 (G_{2j} - x)^2 \right\}$$

Si ottiene anche:  $W(G_2, P_2) \leq W(G_2, P_1)$  (cfr proposizione e la sua prova più avanti)

### k<sup>a</sup> tappa

Si dispone la partizione:  $P_{k-1} = \{A_{(k-1)1}, A_{(k-1)2}, \dots, A_{(k-1)p}\}$

Si scelgono  $p$  nodi  $G_{ki}$  che minimizzano rispettivamente le quantità  $D(y, A_{(k-1)i})$ , cioè: per ogni  $y$ , valore osservato in  $A$ ,  $D(G_{ki}, A_{(k-1)i}) = \sum_{x \in A_{(k-1)i}} (y - x)^2$

Il nodo  $G_{ki}$  è il baricentro di  $A_{(k-1)i}$ . Si ottiene così una nuova successione di nodi:  $\{G_{k1}, G_{k2}, \dots, G_{kp}\}$  di cui  $G_k$  è il baricentro.

Si determina la partizione  $P_k = \{A_{k1}, A_{k2}, \dots, A_{kp}\}$  tale che:

$$A_{ki} = \left\{ x \mid \sum_j (G_{ki} - x)^2 (G_{kj} - x)^2 \right\}$$

Si ottiene ancora:  $W(G_k, P_k) \leq W(G_k, P_{k-1})$ .

Il processo è finito. In effetti, la successione  $W(G_k, P_k)$  è non crescente. Se essa diventa stazionaria, essa converge verso il suo minimo, i nodi e le partizioni essendo invariate per il

fatto che la somma  $\sum_{i=1}^p D(G_i, A_i)$  è positiva e costituita da elementi non crescenti.

### Esempio

Supponiamo che abbiamo osservato i pesi seguenti di 17 individui:

$$A = \{54, 55, 55, 57, 57, 58, 60, 66, 67, 67, 68, 70, 71, 74, 78, 79, 79\}$$

#### 1<sup>a</sup> tappa

Si sceglie

$$* \text{ la partizione } P_0 = \{[50, 60[, [60, 70[, [70, 80]\}$$

$$* \text{ ed i nodi : } G_{11} = 56 ; G_{12} = 65,25 ; G_{13} = 75,17$$

Si calcolano i valori  $(G_{11} - x)^2$  per tutte le  $x$  di  $A$ , poi nello stesso modo tutte le  $(G_{12} - x)^2$  e le  $(G_{13} - x)^2$ . Si associa a  $G_{11}$  i valori di  $x$  che minimizzano le espressioni  $(G_{11} - x)^2$ , siano i numeri : 54, 55, 55, 57, 57, 58, 60.

Dunque  $A_{11} = \{54, 55, 55, 57, 57, 58, 60\}$ .

$$\text{Si ottiene anche : } A_{12} = \{66, 67, 67, 68, 70\}, \text{ poi } A_{13} = \{71, 74, 78, 79, 79\}$$

#### 2<sup>a</sup> tappa

Si scelgono per nodi i baricentri rispettivi di  $A_{1i}$ , per  $i=1,2,3$ .

$$\text{Sia } G_{21} = 56,57 ; G_{22} = 67,75 ; G_{23} = 76,2$$

$$\text{Si ottiene facilmente : } A_{21} = \{54, 55, 55, 57, 57, 58, 60\}.$$

$$A_{22} = \{66, 67, 67, 68, 70, 71\}, \text{ poi } A_{23} = \{74, 78, 79, 79\}$$

#### 3<sup>a</sup> tappa

Si scelgono per nodi i baricentri rispettivi di  $A_{2i}$ , per  $i=1,2,3$ .

$$\text{Sia } G_{31} = 56,57 ; G_{32} = 68,4 ; G_{33} = 77,5$$

$$\text{Si ottiene : } A_{31} = A_{21} ; A_{32} = A_{22} ; A_{33} = A_{23}$$

Il processo dunque converge e l'algoritmo si arresta sull'ultima partizione.

### Nota

Al fine di valutare la qualità della partizione ottenuta, si calcola il rapporto

$$\tau = \frac{\text{Inerzia inter-classe}}{\text{Inerzia totale}}$$

Ora l'inerzia inter-classe di  $A$  per la partizione  $P_3 = \{A_{31}, A_{32}, A_{33}\}$  è :

$$m(G_{3i})(G_{3i} - G)^2 = 47,433 \text{ dove } m(G_{3i}) \text{ è l'effettivo di } x \text{ in } A_{3i} \text{ e } G \text{ è il baricentro di } A.$$

$i=1,2,3$

L'inerzia totale è  $\sum_{x \in A} (G - x)^2 = 49,82$ , siano il tasso eccellente di 0,95.

### 1-3-3 Secondo problema

Si suppone adesso che gli intervalli  $[ \alpha_1, \alpha_2 ]$  et  $[ \beta_1, \beta_2 ]$  sono muniti di partizioni ottimali  $P$  e  $Q$ , rispettivamente, nel senso delle nuvole dinamiche. Siano  $p$  e  $q$  i numeri rispettivi dei sotto intervalli componente  $P$  e  $Q$ . A partire da queste due partizioni, è possibile generare  $2^{p-1}$  e  $2^{q-1}$  partizioni ottenute da unioni iterate di sotto intervalli adiacenti rispettivamente di  $P$  e di  $Q$ <sup>5</sup>

<sup>5</sup> E' sufficiente considerare l'arborescenza di cui  $A_1$  è la radice, poi l'unione o non ad  $A_2$ , che egli stesso sarà o no unito ad  $A_3$ , etc. Vi sono dunque  $2^{p-1}$  branche in questa arborescenza.



Si calcolano le intensità di implicazione rispettive di ogni sotto intervallo unito o non ad un altro della prima partizione su ogni sotto intervallo unito o non ad un altro della seconda, poi i valori delle intensità delle implicazioni reciproche.

Vi sono in totale  $2 \cdot 2^{p-1} \cdot 2^{q-1}$  famiglie di intensità d'implicazione, ciascuna di esse necessitando il calcolo di tutti gli elementi di una partizione di  $[ \alpha, \beta ]$  su tutti gli elementi di una partizione di  $[ \gamma, \delta ]$  e reciprocamente .

Si sceglie come *criterio di ottimizzazione* la media geometrica delle intensità d'implicazione, media associata ad ogni coppia di partizioni di elementi uniti o non definiti induttivamente. Si annotano i due massimi ottenuti (implicazione diretta e la sua reciproca) e si prendono le due partizioni associate dichiarando che l'implicazione della variabile-intervallo a sulla variabile-intervallo b è ottimale quando l'intervallo  $[ \alpha, \beta ]$  ammette la partizione corrispondente al primo massimo e che l'implicazione reciproca ottimale è soddisfatta dalla partizione  $[ \gamma, \delta ]$  corrispondente al secondo massimo.

### Nota

1° Non esiste relazione d'ordine tra  $(a, \bar{b})$  e  $(a, \overline{b \bar{c}})$  conoscendo  $(a, \bar{b})$  e  $(a, \bar{\delta})$  .

In effetti si può avere  $(a, \bar{b}) < (a, \overline{b \bar{c}})$  e  $(a, \bar{\delta}) < (a, \overline{b \bar{c}})$  .

**esempio 1** :  $n = 100$  ;  $n_a = 16$  ;  $n_b = 35$  ;  $n_c = 30$  ;  $n_{a \bar{b}} = 10$  ;  $n_{a \bar{c}} = 8$  ;  $n_{a \overline{(b \bar{c})}} = 2$

Ma si può egualmente avere :  $(a, \bar{b}) > (a, \overline{b \bar{c}})$  e  $(a, \bar{\delta}) > (a, \overline{b \bar{c}})$

**esempio 2** :  $n = 100$  ;  $n_a = 30$  ;  $n_b = 50$  ;  $n_c = 49$  ;  $n_{a \bar{b}} = 15$  ;  $n_{a \bar{c}} = 16$  ;  $n_{a \overline{(b \bar{c})}} = 1$

2° Ancora, non esiste a priori relazione tra  $(a, \bar{\delta})$  ,  $(b, \bar{c})$  e  $(a \bar{b}, \bar{\delta})$

In effetti, si può avere,  $(a, \bar{\delta}) < (a \bar{b}, \bar{\delta})$  e  $(b, \bar{c}) < (a \bar{b}, \bar{\delta})$

**esempio 3** :  $n = 100$  ;  $n_a = 20 = n_b$  ;  $n_c = 35$  ;  $n_{a \bar{c}} = 16 = (b, \bar{c})$  ;  $n_{a \bar{b}, \bar{c}} = 10$

Ma si può avere ugualmente:  $(a, \bar{\delta}) > (a \bar{b}, \bar{\delta})$  et  $(b, \bar{c}) > (a \bar{b}, \bar{\delta})$

**esempio 4** :  $n = 100$  ;  $n_a = 20 = n_b$  ;  $n_c = 48$  ;  $n_{a \bar{c}} = 11 = (b, \bar{c})$  ;  $n_{a \bar{b}, \bar{c}} = 22$

Questo mostra che l'algoritmo di ricerca dell'ottimizzazione delle implicazioni nel corso delle unioni successive deve funzionare "fino alla fine", cioè sino a quando tutte le unioni sono state prodotte e stimate rispetto alla loro potenza implicativa.

### Decrescenza della funzione W

#### Alcune note

Sia A l'insieme dei valori osservati nell'intervallo  $[ \alpha, \beta ]$  e  $L = \{N_1, N_2, \dots, N_p\}$  un insieme di parti di A.  $N_i$  è chiamato  $i^{\text{esimo}}$  nodo di L. Questi nodi sono scelti in modo tale che  $\text{card } N_i$  sia la stessa per ogni i.

Sia  $P = \{A_1, A_2, \dots, A_p\}$  una partizione di A in p classi.

$$\text{Si pone } W(L,P) = \sum_{i=1}^p D(N_i, A_i) = \sum_{i=1}^p \sum_{x \in A_i, y \in N_i} d(x, y)^6$$

$D(N_i, A_i)$  è una sorta di misura di dissomiglianza tra il nodo  $N_i$  e la classe  $A_i$ .

Il problema delle nuvole dinamiche porta a minimizzare  $W(L, P)$  con la costruzione di un insieme conveniente di  $p$  nodi in  $L^*$  e di una partizione  $P^*$  in  $p$  classi.

Algoritmo di nuvole dinamiche

I nodi  $L_0 = \{N_{01}, N_{02}, \dots, N_{0p}\}$  sono dati (o scelti arbitrariamente).

Si definisce la partizione  $P_1 = \{A_{11}, A_{12}, \dots, A_{1p}\}$  che si deduce da:

$$A_{1i} = \left\{ x \in A / \min_j d(N_{0i}, x) \leq \min_j d(N_{0j}, x) \right\}.$$

In caso di uguaglianza, si assegna  $x$  alla classe di indice più piccolo.

Si pone  $P_1 = f(L_0)$ .

Si costruiscono allora i nodi di  $L_1 = \{N_{11}, N_{12}, \dots, N_{1p}\}$  con il procedimento :

$$N_{1k} = \left\{ x \in A / d(x, A_k) = \inf_j d(x, A_{1j}) \right\}$$

Si scrive  $L_1 = g(P_1)$  e si itera l'algoritmo.

**Proposizione :**

$W(L,P)$  decresce ad ogni iterazione, i.e. :

- a)  $P, W(L, f(L)) \leq W(L, P)$
- b)  $L, W(g(P), P) \leq W(L, P)$

-----

**Prova :**

b) Sia data una qualunque partizione nel corso dell'algoritmo  $P = \{A_1, A_2, \dots, A_p\}$  e sia  $f(L) = Q = \{B_1, B_2, \dots, B_p\}$  la partizione ottenuta con l'aiuto dei nodi  $L = \{N_1, N_2, \dots, N_p\}$  definita a partire da  $P$ .

$$\text{Allora } W(L,P) = \sum_{i=1}^p D(N_i, A_i) = \sum_{i=1}^p \sum_{x \in A_i, y \in N_i} d(x, y) = \sum_{j=1}^p \sum_{x \in A_j, y \in N_j} d(x, y)$$

$$\text{e } W(L,Q) = \sum_{j=1}^p \sum_{x \in A_j, y \in B_j} d(x, y)$$

Sia  $x \in A_i$ . Per ogni  $i$  e per ogni  $x$  tale che  $x \in A_i \cap B_i$ , allora  $x$  ha lo stesso contributo alle somme  $W(L,P)$  e  $W(L,Q)$ . Di contro, se  $x \in A_i \cap B_j$ , allora  $d(N_j, x) < d(N_i, x)$

Perché per costruzione:  $B_j = \{x \in A / \min_i d(N_j, x) \leq \min_i d(N_i, x)\}$

In seguito, per ogni  $x \in A_i$ , i contributi di  $x$  a  $W(L, Q)$  sono inferiori o uguali a quelli di  $W(L, P)$ , sia ancora  $P, W(L, f(L)) \leq W(L, P)$ .

---

<sup>6</sup>  $d(x,y)$  può essere uguale a  $(x-y)^2$  come è stato scelto nel primo problema.

c) Sia  $g(P) = \{O_1, O_2, \dots, O_p\}$  un insieme di  $p$  nodi ottenuti a seguito della partizione  $P = \{A_1, A_2, \dots, A_p\}$ , essa stessa derivata da  $L = \{N_1, N_2, \dots, N_p\}$

e definita così:

$$k, O_k = \left\{ x \in A / d(x, A_k) = \inf_j d(x, A_j) \right\}.$$

Quindi,  $j = \arg \min_{x \in N_j} d(x, A_j)$  dove  $L, W(g(P), P) = W(L, P)$

#### 1-4 L'implicazione-inclusione

Due ragioni ci hanno condotto a migliorare il modello realizzato dall'intensità di implicazione.

- quando la taglia dei campioni trattati, ed in particolare quello di  $E$ , cresce (dell'ordine del migliaio o più), l'intensità  $(a, \bar{b})$  ha tendenza a non essere più sufficientemente discriminante perché i valori possono essere vicini ad 1, allorché l'inclusione che cerca di modellizzare la qualità, è lontana dall'essere soddisfatta (fenomeno segnalato da A. Bodin che tratta grandi popolazioni di allievi attraverso indagini internazionali);
- il modello della quasi-implicazione precedente mantiene essenzialmente la misura della forza della regola  $a \Rightarrow b$ . Ora, la presa di coscienza di una potente concomitanza di  $b \Rightarrow \text{non } a$  (contrapposto di implicazione) è indispensabile per rinforzare l'affermazione di una buona qualità della relazione quasi-implicativa di  $a$  su  $b$ . Nello stesso tempo, essa potrebbe permettere di correggere la difficoltà evocata prima (se  $A$  e  $B$  sono piccoli per rapporto ad  $E$ , i loro complementari saranno importanti e reciprocamente).
- La soluzione che portiamo utilizza contemporaneamente l'intensità d'implicazione ed un altro indice che rende conto della qualità dell'informazione fornita dalla debolezza relativa delle istanze che contraddicono la regola e la sua contrapposta. E' dunque ad un concetto di entropia di Shannon che noi facciamo riferimento:

$$H(b/a=1) = - \frac{n_{a \bar{b}}}{n_a} \log_2 \frac{n_{a \bar{b}}}{n_a} - \frac{n_{a b}}{n_a} \log_2 \frac{n_{a b}}{n_a},$$

Entropia condizionale relativa ai casi  $(a \text{ e } b)$  e  $(a \text{ e non } b)$  quando  $a$  è realizzata

$$H(\bar{a}/\bar{b}=1) = - \frac{n_{a \bar{b}}}{n_{\bar{b}}} \log_2 \frac{n_{a \bar{b}}}{n_{\bar{b}}} - \frac{n_{\bar{a} \bar{b}}}{n_{\bar{b}}} \log_2 \frac{n_{\bar{a} \bar{b}}}{n_{\bar{b}}}$$

Entropia condizionale relativa ai casi  $(\text{non } a \text{ e non } b)$  e  $(a \text{ e non } b)$  quando  $\text{non } b$  è realizzato.

Queste entropie dovranno dunque essere simultaneamente piccole se si spera di disporre di un buon criterio di inclusione di  $A$  in  $B$ . Tuttavia con una media aritmetica debole di queste due entropie, noi disponiamo di un buon criterio d'inclusione, che bisognerebbe adesso adattare al modello atteso nelle differenti situazioni cardinali.

Infatti, al fine di rispettare i numeri delle rispettive istanze di  $a$  in rapporto a  $b$  e di  $\text{non } b$  in rapporto a  $\text{non } a$ , l'inclusione di  $A$  in  $B$  avrà intanto più senso che sarebbero deboli nell'osservazione le due funzioni  $h_1$  e  $h_2$  definite rispettivamente a partire da  $H(b/a=1)$  e  $H(\bar{a}/\bar{b}=1)$ , da :

$$h_1(a,b) = \frac{1}{2} \left( -\frac{n_a \bar{b}}{n_a} \log_2 \frac{n_a \bar{b}}{n_a} - \frac{n_a \bar{b}}{n_a} \log_2 \frac{n_a \bar{b}}{n_a} \right) \mathbf{1}_{0; \frac{n_a}{2n}}(t) +$$

$$\frac{1}{2} \left( 2 + \frac{n_a \bar{b}}{n_a} \log_2 \frac{n_a \bar{b}}{n_a} + \frac{n_a \bar{b}}{n_a} \log_2 \frac{n_a \bar{b}}{n_a} \right) \mathbf{1}_{\frac{n_a}{2n}; \frac{n_a}{n}}(t)$$

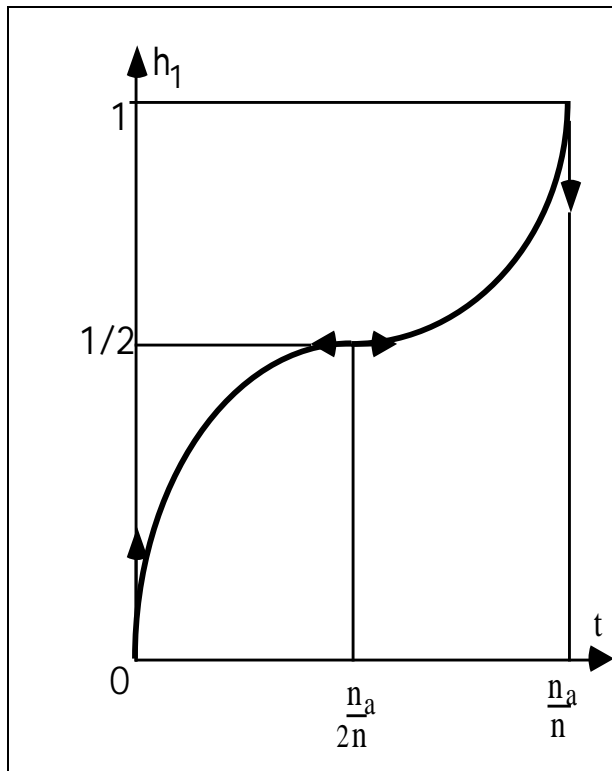
$$h_2(a,b) = \frac{1}{2} \left( -\frac{n_a \bar{b}}{n_b} \log_2 \frac{n_a \bar{b}}{n_b} - \frac{n_a \bar{b}}{n_b} \log_2 \frac{n_a \bar{b}}{n_b} \right) \mathbf{1}_{0; \frac{n_b}{n}}(t) +$$

$$\frac{1}{2} \left( 2 + \frac{n_a \bar{b}}{n_b} \log_2 \frac{n_a \bar{b}}{n_b} + \frac{n_a \bar{b}}{n_b} \log_2 \frac{n_a \bar{b}}{n_b} \right) \mathbf{1}_{\frac{n_b}{2n}; \frac{n_b}{n}}(t)$$

dove si osserva  $t = \frac{n_a \bar{b}}{n}$ , e per esempio,  $\mathbf{1}_{0; \frac{n_a}{2n}}$  la funzione indicatrice sull'intervallo

$0; \frac{n_a}{2n}$  de la variabile  $t$ , frequenza dei contro-esempi, per  $n_a$  e  $n_b$  fissati.

Rappresentiamo per esempio la funzione  $h_1$  di  $t$  :



Si constata che questa rappresentazione di funzione continua e derivabile di  $t$  traduce le proprietà attese dal criterio di inclusione:

- \* "reazione" rapida ai primi contro-esempi,
- \* "rallentamento" del rifiuto dell'inclusione nell'intorno del punto di equilibrio  $\frac{n_a}{2n}$ ,
- \* rifiuto sempre più accentuato al di là di  $\frac{n_a}{2n}$  ciò che non assicurava l'intensità di implicazione.

In definitiva, consideriamo:

### Definizione 6

Come l'indice di implicazione, il numero:

$$i(a,b) = 1 - \frac{1}{2}(h_1(a,b) + h_2(a,b))$$

che integra l'informazione consegnata dalla realizzazione di un debole numero di contro-esempi di una parte alla regola  $a \Rightarrow b$ , dall'altra parte, alla regola non  $b \Rightarrow$  non  $a$  come intensità d'implicazione-inclusione (o intensità entropica) il numero:

$$(a, b) \neq (i(a, b), (a, \bar{b}))^{\frac{1}{2}}$$

che integra contemporaneamente lo stupore statistico e la qualità inclusiva.

### Esempio

	b	$\bar{b}$	marge	L'intensità d'implicazione è 0,9999 ( $q(a, \bar{b}) = -3,65$ )
a	200	400	600	L'entropia media dell'esperienza è uguale a : 0,406 ( $h_1 = 0,541$ ,
$\bar{a}$	600	2800	3400	$h_2 = 0,272$ ). Il valore del coefficiente moderatore è: $i(a, b) = 0,594$
marge	800	3200	4000	Dunque $(a, b) \neq 0,77$

Così, l'entropia "modera" l'intensità d'implicazione nei casi dove giustamente l'inclusione è mediocre, utilizzando la qualità della contrapposta, come sopra. Se ciascuno dei cardinali della tabella precedente è diviso per 100, il coefficiente  $i(a, b)$  non cambia allora che l'intensità che diviene 0,602.

### **Nota**

La corrispondenza tra  $(a, \bar{b})$  e  $i(a, b)$  non è monotona come mostra il seguente esempio:

	b	$\bar{b}$	marge	L'intensità d'implicazione è superiore alla precedente: $q(a, \bar{b}) = -4,38$
a	250	450	700	L'entropia media dell'esperienza è uguale a : 0,411 ( $h_1 = 0,530$ ,
$\bar{a}$	600	2750	3350	$h_2 = 0,2929$ ). Il valore del coefficiente moderatore è: $i(a, b) = 0,588$
marge	850	3200	4050	Dunque $(a, b) \neq 0,767$

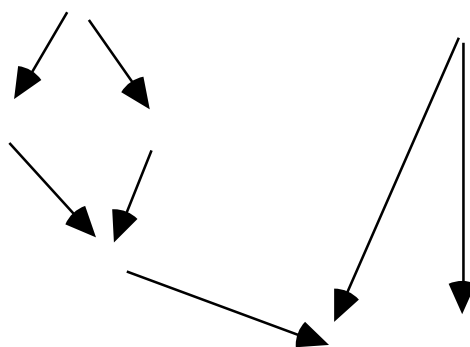
Così, allorché  $(a, \bar{b})$  è cresciuto dal 1° al 2° esempio,  $i(a, b)$  è diminuito. Invece, la situazione contraria è la più frequente.

### **1-5 Grafo d'implicazione**

La relazione definita dall'implicazione statistica, se essa è riflessiva e non simmetrica, non è transitiva in maniera evidente. Ora vogliamo che essa modella la relazione d'ordine parziale tra due variabili (i successi nel nostro esempio iniziale). Per convenzione, se  $a \Rightarrow b$  e se  $b \Rightarrow c$ , accetteremo la chiusura transitiva  $a \Rightarrow c$  solamente se  $(a, \bar{c}) > 0,5$ , cioè se la relazione implicativa di a su c è migliore che la neutralità.

Per esempio, supponiamo che tra le 7 variabili a, b, c, d, e, ed f esistano, alla soglia superiore a 0.5, le relazioni seguenti  $e \Rightarrow c, a, f, b$ ;  $c \Rightarrow a, f$ ;  $b \Rightarrow a, f$ ;  $g \Rightarrow d, f$ ;  $a \Rightarrow f$ .

Si potrà allora tradurre questo insieme di relazioni con il grafo seguente:



## § 2 Implicazione tra regole e meta-regole

Una implicazione tra classi di variabili prende il suo significato solo a condizione che all'interno di ogni classe di variabili di cui si esamina la relazione con le altre, esiste una certa "coesione" tra le variabili che la costituiscono. Si spera così che il "flusso" implicativo di una classe **A** su una classe **B** sia alimentato da un "flusso" interno ad **A** e alimenta un "flusso" interno a **B** (questa parola "flusso" è scelta per la sua connotazione metaforica idraulica o termodinamica). Per questo, il concetto d'entropia  $H$  che permette di spiegare il disordine tra due variabili, definiamo la coesione tra due variabili:

### Definizione 7

La coesione della classe  $(a, b)$  è il numero  $c(a, b)$  tale che:

se  $p = \max ( (a, \bar{b}), (b, \bar{a}) )$  et  $H = -p \log_2 p - (1-p) \log_2 (1-p)$

allora  $c(a,b) = \sqrt{1 - H^2}$

se  $p=1$  allora  $c(a,b) = 1$

se  $p \leq 0,5$  allora  $c(a,b) = 0$

### Definizione 8

La coesione della classe  $\underline{A} = (a_1, \dots, a_r)$  è allora definita per estensione:

$$C(\underline{A}) = \frac{2}{r(r-1)} \sum_{\substack{i \in \{1, \dots, r-1\} \\ j \in \{2, \dots, r\}, j > i}} c(a_i, a_j)$$

**E' la media geometria delle coesioni delle classi a due elementi.**

Infine possiamo modellizzare l'implicazione statistica di una classe di variabili su di un'altra classe esigendo dal modello che esso integri le informazioni seguenti:

- le coesioni rispettive di due classi,
- una intensità d'implicazione estrema di elementi di una classe sugli elementi dell'altra,
- i cardinali rispettivi di due classi.

Ciascuna di queste informazioni accredita l'indice che presentiamo così:

- l'indice cresce con le coesioni di ogni classe e s'annulla quando la coesione di una di esse è nulla,
- l'indice cresce con il legame estremo (minimale se si mira ad un grado di esigenza elevato, massimale se si cerca una flessibilità realista),
- l'indice decresce con i cardinali delle classi, avuto riguardo della presa di coscienza del legame massimale.

In seguito, notando con **A** e **B** classi di variabili di elementi generici  $a_i$  e  $b_j$ , poi  $C(\underline{A})$  e  $C(\underline{B})$  loro coesioni rispettive, l'intensità d'implicazione di **A** su **B** è data:

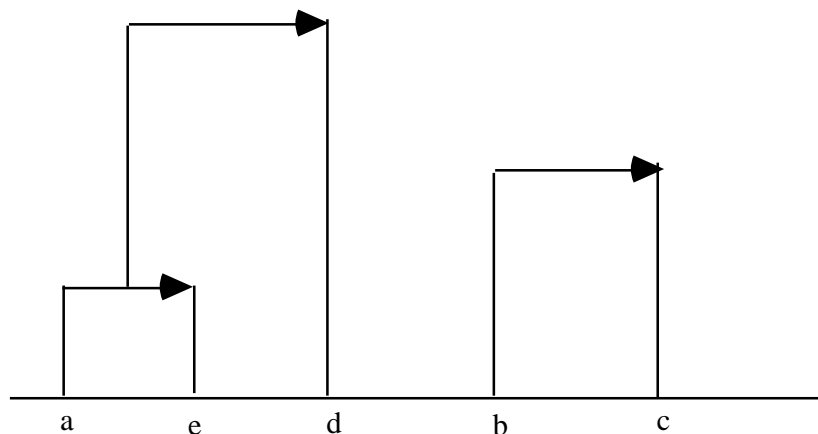
**Definizione 9**

L'intensità d'implicazione di **A** su **B** è:

$$I(\underline{A}, \underline{B}) = \sup_{i \in \{1, \dots, r\}, j \in \{1, \dots, s\}} (a_i, \bar{b}_j) \cdot [C(\underline{A}) \cdot C(\underline{B})]^{1/2}$$

Si potrà constatare che questo indice soddisfa gli obblighi semantici dichiarati prima.

Definendo a partire da questo indice un metodo di classificazione discendente classica da un criterio di coesione decrescente, si otterranno per esempio degli alberi come questo:



**§ 3 Livelli significativi di una gerarchia implicativa**

Essendo data la molteplicità dei livelli di formazione delle classi, è indispensabile svincolare quelle che sono più pertinenti in rapporto all'intenzione classificatoria del ricercatore e avendo riguardo dei criteri scelti. Procediamo (cfr Lermann I. C., 1981) allora comparabilmente a quella adottata per prima da Lermann relativamente alla gerarchia di similarità, ma ricondizionandone il suo approccio.

**3.1. Preordine coesivo**

Consideriamo l'insieme  $V$  delle variabili  $\{a_1, a_2, \dots, a_m\}$  e l'insieme delle coppie  $(a, b)$  di  $V \times V$  tali che  $a \leq b$ . Esistono  $m(m-1)$  di queste coppie alle quali si sono associate le loro coesioni  $c(a, b)$  rispettive.

**Definizione 10:**

**Si chiama preordine iniziale e globale coesivo su  $V \times V$  (o preordinamento), il preordine  $\Omega$  indotto dall'applicazione coesione  $c$  su  $V \times V$ .**

Sia  $G(\ )$  il suo grafo in  $V \times V$ . Dai §1 e §2 che precedono, segue che:

- la classe del perordine corrispondente a  $c=0$  contiene tutte le coppie tali che  $(a, \bar{b}) \in \Omega$ ,
- $n_a \leq n_b$  allora  $c(b,a) \leq c(a,b)$ .

Notiamo, di contro, che  $c(a,b) \leq c(c,d)$  non si ha necessariamente  $c(b,a) \leq c(d,c)$  o  $c(b,a) \leq c(d,c)$ .

**3.2. Determinazione dei livelli significativi**

Mettiamoci ad un livello qualunque  $k$  della gerarchia. A questo livello, si forma una classe di  $m_i$  variabili ( $2 \leq m_i \leq m$ ) di cui la coesione è meno buona di quella delle classi anteriormente formate, conformemente all'algoritmo preso in considerazione, e migliore di quelle delle classi che verranno.

Sia  $\pi_k$  la partizione su  $V$  definita a questo livello costituita da classi che si sono già formate e, evidentemente, dei singletons non ancora associati.  $\pi_k$  è più fine di  $\pi_{k+1}$ .

Sia  $S_k$  l'insieme delle coppie separate a questo livello e  $R_k$  l'insieme delle coppie che vi sono unite per la prima volta, essendo inteso che si dirà che la coppia  $(a, b)$  è unita se  $a$  e  $b$  appartengono alla stessa classe del tipo  $(\dots(\dots a, \dots) \dots b) \dots$ .

L'insieme  $G(\ ) \cap [S_k \times R_k]$  è costituito da coppie di coppie che al livello  $k$  rispettano il preordine iniziale. Per esempio, se si ha  $c(e,f) < c(a,b)$  (dunque  $((e,f), (a,b)) \in G(\ )$ ) e se al livello  $k$ ,  $e$  ed  $f$  sono separati allorché  $a$  e  $b$  si riuniscono nella classe che si forma, la coppia  $((e,f), (a,b))$  appartiene a  $G(\ ) \cap [S_k \times R_k]$ .

Come è stato fatto per il cardinale  $A \in \bar{B}$ , associamo (cfr. LERMAN I.C. 1981) al cardinale di  $G(\ ) \cap [S_k \times R_k]$  l'indice aleatorio  $\text{card}[G(\ *) \cap [S_k \times R_k]]$  dove  $\ast$  è un preordinamento aleatorio nell'insieme, munito di una probabilità uniforme, di tutte i preordinamenti dello stesso tipo cardinale di  $\ast$ . Questo indice ha come speranza  $1/2 \text{card}[S_k \times R_k]$  e per varianza  $\text{card}[S_k \times R_k] \text{card}[G(\ )]$ .

Sia  $s(\ast, k)$  l'indice centrato ridotto ottenuto:

$$\frac{(\text{card}[G(\ast) \cap [S_k \times R_k]] - 1/2 \text{card}[S_k \times R_k])}{(\text{card}[S_k \times R_k] \text{card}[G(\ast)])^{1/2}}$$

**Definizione 11**

**Si chiama nodo significativo ogni nodo corrispondente ad un massimo locale di  $s(\Omega, k)$  nel corso della costituzione della gerarchia implicativa. Diremo in questo caso che la partizione  $\pi_k$  è in risonanza parziale con  $\Omega$ .**

**Se, inoltre,  $G(\Omega) \cap [S_{\pi_k} \times R_{\pi_k}] = S_{\pi_k} \times R_{\pi_k}$ , noi diremo che la partizione  $\pi_k$  è in risonanza totale con  $\Omega$ .**



## § 4 Contribuzione dei soggetti e delle variabili supplementari

Introduciamo la nozione di variabile supplementare in analisi implicativa analoga alla stessa nozione definita in analisi fattoriale, cioè variabile estrinseca, descrittore per esempio, che non interviene direttamente nei legami espressi dalla classificazione tra le variabili dette principali di  $V$ , dunque non intervenendo nella struttura di questo insieme sotto la forma grafo o gerarchia. Per esempio, una variabile supplementare potrà rappresentare una categoria di soggetti (età, sesso, categoria socio-professionale, etc.).

Ad un livello qualunque della gerarchia si forma una classe  $C$  di coesione non nulla. Il nostro obiettivo, particolarmente nel caso di un nodo significativo, è di definire un criterio che permetta di identificare uno o dei soggetti, poi le categorie del soggetto, o ogni altra variabile supplementare (età, sesso, categoria socio-professionale, etc.), che contribuisca inoltre alla costituzione della classe. Il comportamento di questi soggetti sarà così in armonia con il comportamento statistico all'origine della classe. Un approccio comparabile è stato fatto congiuntamente per studiare il contributo delle variabili supplementari alla costituzione di un arco o di un cammino del grafo implicativo.

### 4.1. Potenza implicativa d'una classe

Mettiamoci al livello  $k$  della gerarchia dove si riuniranno, per formare  $C$ , due classi  $A$  e  $B$  tali che  $A \Rightarrow B$  nel senso del § 2.

#### Definizione 12

**La coppia  $(a, b)$  tale che:  $\forall i \in A, \forall j \in B \quad \varphi(a, \bar{b}) \geq \varphi(i, \bar{j})$  è chiamata coppia generica di  $C$ . E' questa coppia, generalmente unica, che interviene nel sup. nel calcolo dell'implicazione di  $A$  su  $B$ . Il numero  $\varphi(a, \bar{b})$  è chiamato implicazione generica di  $C$ .**

Ma, in ogni sotto classe  $C$ , esiste ugualmente una coppia generica. Precisamente, se  $C$  è costituita da  $g$  ( $g \geq k$ ) sotto classi ( $C$  compreso), vi sono  $g$  coppie generiche all'origine di  $C$  e  $g$  intensità massimali d'implicazione  $\varphi_1, \varphi_2, \dots, \varphi_g$  che le corrispondono.

#### Definizione 13

**Il vettore  $(\varphi_1, \varphi_2, \dots, \varphi_g)$ , elemento di  $[0,1]^g$ , è chiamato vettore potenza implicativa di  $C$ , traducendo una forza implicativa interna a  $C$ .**

### 4.2. Potenza implicativa di un soggetto su una classe e distanza a questa classe

Un soggetto  $x$  qualunque rispetta o no l'implicazione della coppia generica di una classe. Associando logica formale e considerazione semantica, porremo, in funzione dei valori presi da  $a$  e  $b$  in  $x$ :  $x(a, \bar{b})=1$  se  $a=1$  dove  $0$  e  $b=1$ ;  $x(a, \bar{b})=0$  se  $a=1$  e  $b=0$ ;  $x(a, \bar{b})=p$  se  $a=b=0$  con  $p \in ]0,1]$ . Più spesso, scegliamo  $p=0.5$ , valore neutro.

Così, a  $x$ , possiamo associare  $n$  numeri  $x_{x,1}, x_{x,2}, \dots, x_{x,g}$  corrispondente ai valori presi in  $x$  dalle  $g$  implicazioni generiche della classe  $C$ .

#### Definizione 14

Il vettore  $\varphi_{x,1}, \varphi_{x,2}, \dots, \varphi_{x,g}$ , elemento di  $[0,1]^g$ , è chiamato **vettore potenza implicativa di x**.  
 Il soggetto  $x_i$ , può essere fittizio, per cui tutte le componenti del vettore potenza sono uguali ad 1 è chiamato **soggetto ideale teorico di C**.

In queste condizioni, si può munire lo spazio delle potenze  $[0,1]^g$ , di una metrica del tipo  $\chi^2$  affine d'accentuare gli effetti delle forti implicazioni generiche.

#### Definizione 15

Si chiama **distanza implicativa di un soggetto x dalla classe C il numero:**

$$d(x,C) = \frac{1}{g} \sum_{i=1}^g \frac{[1 - \varphi_{x,i}]^2}{1 - \varphi_{x,i}} \quad \frac{1}{2}$$

Questo numero non è altro che la distanza detta del  $\chi^2$  tra le due distribuzioni  $\{1 - \varphi_{x,i}\}_i$  e  $\{1 - \varphi_{x,i}\}_i$  che esprimono gli scarti tra le implicazioni generiche empiriche e l'implicazione stretta. Se per un  $i$ ,  $\varphi_{x,i} = 1$ , noi porremo, per convenzione,  $\frac{1}{1 - \varphi_{x,i}} = 1$ . Questa convenzione non si fa contro natura, poiché in questo caso, l'implicazione generica è massimale e significativa di un eccellente legame implicativo tra i due suoi termini, verificato da tutti i soggetti  $x$  di E. Così, se il denominatore si annulla, lo è anche il numeratore, e si potrà attribuire il valore 0 al quoziente.

#### 4.3. Contribuzione di un soggetto e di una variabile supplementare ad una classe

Definiremo a partire della "distorsione" del soggetto considerato per rapporto al soggetto ideale teorico, ricordando che possono esistere soggetti reali la cui distanza dalla classe C sia inferiore alla distanza a questa stessa classe del soggetto ideale teorico. Il contributo di una categoria del soggetto o di una variabile supplementare G ne sarà dedotto.

#### Definizione 16

Il contributo di x a C è:  $(x, C) = \frac{d(x_t, C)}{d(x, C)}$

E quello di G è:  $(G, C) = \frac{1}{\text{card}G} \sum_{x \in G} (x, G)$

Questi contributi possono essere infiniti (per delle configurazioni contenenti degli  $x$  a distanza nulla da C) ma, in particolare, superiori ad 1 per certi soggetti.

Alfine di dare al ricercatore lo strumento di sapere o di verificare rapidamente se tale categoria dei soggetti che lo interessano è statisticamente determinata nella costituzione di una classe implicativa, un algoritmo è stato elaborato appoggiandosi sulle due nozioni seguenti: gruppo ottimale e categoria determinata.

### **Definizione 17**

**Sia E la popolazione studiata. Un gruppo ottimale di una classe implicativa C, chiamata GO(C), è il sotto insieme di E che accorda a questa classe un contributo più grande del suo complementare e che forma con questo una partizione in due classi che rendono massima la varianza inter-classe della serie statistica dei contributi individuali. Una tale partizione è detta significativa.**

L'esistenza di questo gruppo ottimale è dimostrato in [GRAS R. e RATSIMBA-RAJOHN H. 1997]. Le proprietà utilizzate sono anche quelle che lo sono per stabilire l'algoritmo sul quale si basano i moduli dei programmi informatici che costruiscono automaticamente nello C.H.I.C. ogni sotto gruppo ottimale.

Consideriamo una partizione  $\{G_i\}_i$  di E,  $X_i$  una parte aleatoria di E avendo lo stesso cardinale di  $G_i$ , e  $Z_i$  la variabile aleatoria  $\text{Card}(X_i \cap \text{GO}(C))$ .  $Z_i$  segue una legge binomiale di parametri:  $\text{card } G_i$  e  $\text{card } \text{GO}(C) / \text{card } E$ .

### **Definizione 18**

**Si chiama la categoria più contributiva alla costituzione della classe implicativa C, la categoria che minimizza l'insieme  $\{p_i\}_i$  delle probabilità  $p_i$  tali che:**

$$\forall i, p_i = \text{Prob} [\text{card } G_i \cap \text{GO}(C) < Z_i]$$

**Una categoria  $G_0$  è detta determinata alla soglia  $\alpha$  se la probabilità associata  $p_0$  è inferiore ad  $\alpha$ .**

Così, il significato di una classe essendo data dall'esperto, associerà la sotto popolazione più portatrice di questo significato. Questo approccio è comparabile a quello di I.-C.Lermann per l'analisi delle similarità, ma per mezzo di una modellizzazione e di concetti differenti.

**In conclusione**, le applicazioni del metodo sono fin d'ora pieni di buoni risultati e non solamente nella disciplina dove è stato concepito (la didattica delle matematiche) ma anche in altri domini dell'Educazione, in psicologia, in sociologia, in biologia, etc. Le analisi beneficiano del software C.H.I.C., sviluppato sotto Windows 95 da R. Couturier, software che permette, con una certa convivialità, tutti i trattamenti dei metodi evocati in questo articolo. Il suo sviluppo segue regolarmente tutte le novità avanzate dalla teoria dell'implicazione statistica.

## Bibliografia

- [AG ALMOULOU S., 1992] - L'ordinateur, outil d'aide à l'apprentissage de la démonstration et traitement de données didactiques, Thèse de doctorat de l'Université de Rennes I.
- [AMARGER S., DUBOIS D., PRADE H., 1991] - Imprecise quantifiers and conditional probabilities - in *Symbolic and quantitative approaches to uncertainty* (R. KRUSE, P. SIEGEL), Springer-Verlag, 33-37.
- [BAILLEUL M., 1994] - Analyse statistique implicative: variables modales et contribution des sujets. Application à la modélisation de l'enseignant dans le système didactique, *Thèse de l'Université de Rennes I, juin 1994.*
- [BAILLEUL M. et GRAS R., 1995] - *L'implication statistique entre variables modales, Mathématique, Informatique et Sciences Humaines, E.H.E.S.S. Paris, n°128*
- [BODIN A., 1997] - Modèles sous-jacents à l'analyse implicative et outils complémentaires. *Prépublication IRMAR. n°97-32*
- [BODIN A. et GRAS R., 1999] : Analyse du préquestionnaire enseignants avant EVAPM-Terminales, *Bulletin n° 425 de l'Association des Professeurs de Mathématiques de l'Enseignement Public, 772-786, Paris*
- [COUTURIER R. et GRAS R., 1999] : Introduction de variables supplémentaires dans une hiérarchie de classes et application à CHIC, Actes des 7èmes Rencontres de la Société Francophone de Classification, 87-92, Nancy, 15-17 septembre 1999
- [DIDAY E., 1972] Nouvelles méthodes et nouveaux concepts en classification automatique et reconnaissance des formes, thèse de doctorat d'état, Université de Paris VI
- [GANASCIA J.G., 1991] - CHARADE : Apprentissages de bases de connaissances dans "Induction symbolique -numérique à partir de données", Ed. KODRATOFF et DIDAY, CEPADUES, 1991.
- [GRAS R., 1979] - Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques, *Thèse d'Etat, Université de Rennes I.*
- [GRAS R. et RATSIMBA-RAJOHN H. 1996] - Analyse non symétrique de données par l'implication statistique. *RAIRO-Recherche Opérationnelle, 30-1, AFCET, Paris.*
- [GRAS R. et PETER P. 1996] - Structuration sets with implication intensity, *Proceedings of the International Conference on Ordinal and Symbolic Data Analysis - OSDA 95, E.Diday, Y.Chevallier, Otto Opitz, Eds., Springer, Paris*
- [GRAS R. et als 1996] - *L'implication Statistique*, Ouvrage de 320 pages dans la Collection Associée à "Recherches en Didactique des Mathématiques", La Pensée Sauvage, Grenoble
- [GRAS R., BODIN A., COUTURIER R., GUILLET F., 1996] - Apprentissage automatique et implication : mise en oeuvre sur un espace d'apprentissage en ressources humaines et Analyse d'une épreuve de concours par la méthode implicative. Présentation interactive. Communications lors des "Quatrièmes Journées de la Société Francophone de classification, Vannes 19-20 Septembre 1996
- [GRAS R., BRIAND H., PETER P., PHILIPPE J., 1997] - *Implicative statistical analysis, Proceedings of International Congress I.F.C.S., 96, Kobé, Springer-Verlag, Tokyo*

- [GRAS R. 1997] : *Metodologia d'analisi di indagini*, in *Quaderni di Ricerca in Didattica*, Palerme n°7, 99-109
- [GRAS R. et PETER P. 1999] : *From a cognitive complexity problem to an implicative model*, *Actes de l'Intensive Programme Socrates/Erasmus 1998/1999, University of Cyprus, in A multidimensional approach to learning in mathematics and sciences, A.Gagatsis Ed., Intercollege Press Cyprus, 491-500, Nicosia*
- [LAGRANGE J.B., 1998] - Analyse implicative d'un ensemble de variables numériques ; application au traitement d'un questionnaire à réponses modales ordonnées, *Revue de Statistique Appliquée., I.H.P. Paris*
- [LARHER A., 1991] - Implication statistique et applications à l'analyse de démarches de preuve mathématique, *Thèse de l'Université de Rennes I.*
- [LERMAN I.C., 1970] - *Les bases de la classification hiérarchique*, chap. 1, Gauthier-Villars, Paris.
- [LERMAN I.C., 1981] - Classification et analyse ordinaire des données, Dunod, 1981.
- [LERMAN I.C., GRAS R., ROSTAM H., 1981] - Elaboration et évaluation d'un indice d'implication pour des données binaires, I et II, *Mathématiques et Sciences Humaines n° 74, p 5-35 et n° 75, p 5-47.*
- [LERMAN I.C., 1998, 1995] - *Rôle de l'inférence statistique dans une approche de l'analyse classificatoire des données*, *Actes du Colloque "Méthodes d'analyses statistiques multidimensionnelles en didactique des mathématiques"*, I.U.F.M. de Caen 27-29 Janvier 1995, Ed. R.Gras, IRMAR, Rennes I
- [LOEVINGER J. 1947] - A systematic approach to the construction and evaluation of tests of abilities, *Psychological Monographs, 61, n° 4*
- [PEARL J. 1988] - Probabilistic Reasoning in intelligent systems, *San Mateo, CA, Morgan Kaufmann.*
- [RATSIMBA-RAJOHN H., 1992] - Contribution à l'étude de la hiérarchie implicative. Application à l'analyse de la gestion didactique des phénomènes d'ostension et de contradictions. *Thèse de doctorat de l'Université de Rennes I.*
- [TOTOHASINA A., 1992] - Méthode implicative en analyse de données et application à l'analyse de conceptions d'étudiants sur la notion de probabilité conditionnelle. *Thèse de doctorat de l'Université de Rennes I.*