Autour de questions méthodologiques soulevées par la construction, le traitement et l'analyse des données utiles à la recherche en sciences humaines et sociales.

Quelques apports de l'Analyse Statistique Implicative : de l'exploratoire au confirmatoire

Jean-Claude Régnier

UMR 5191 ICAR Université Lyon2

icar
UMR 5191
Interactions, Corpus, Apprentissages, Représentations

L ’implication statistique (1)

Un cadre théorique d ’analyse exploratoire des données fondée sur l’analyse statistique implicative

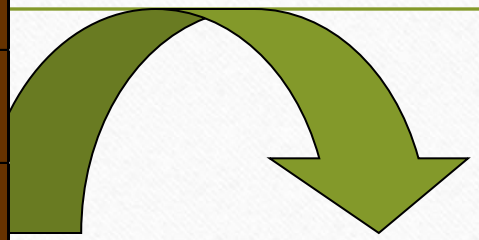
Jean-Claude Régnier

UMR 5191 ICAR Université Lyon2

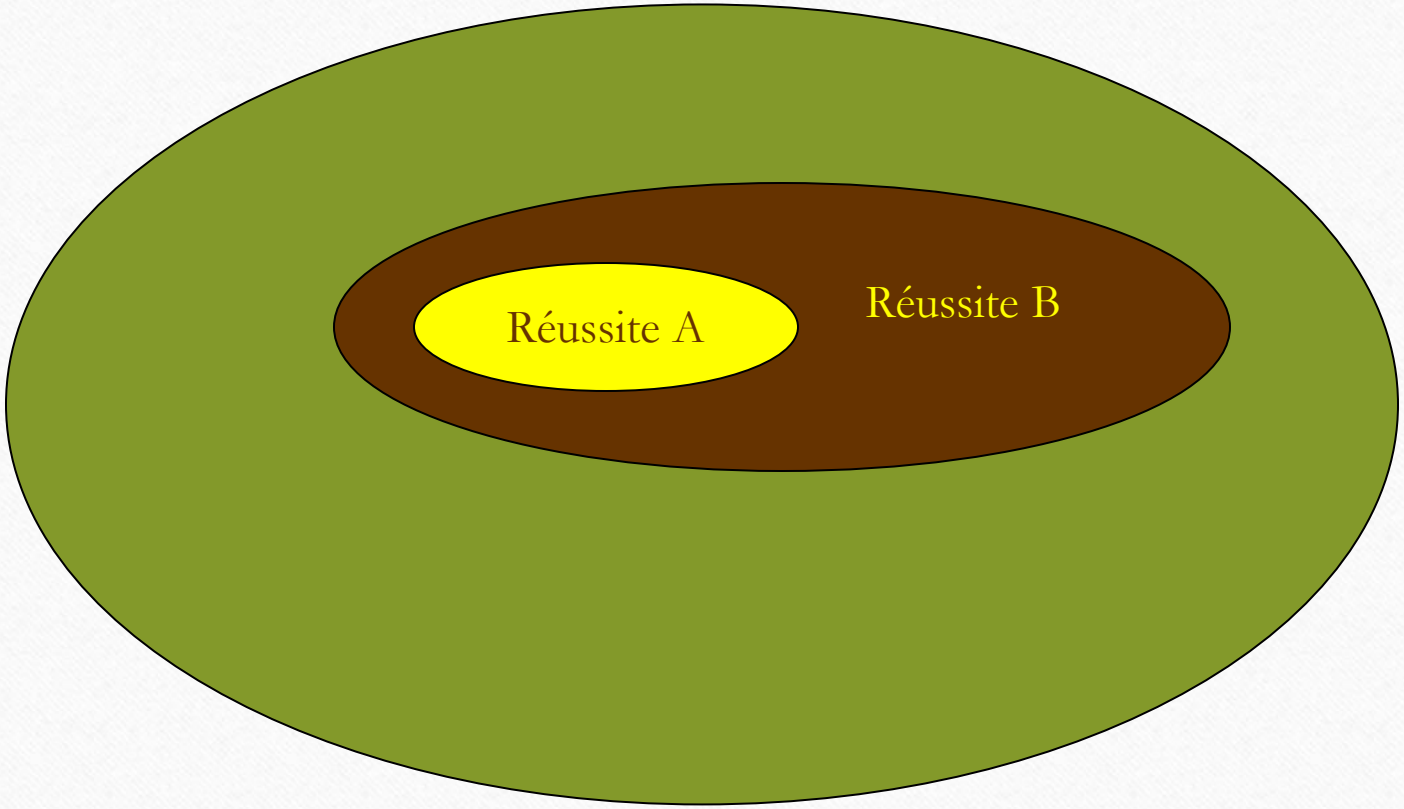
Une question historique dans le développement de l'A.S.I. sur laquelle se sont basés les travaux de Régis Gras (1933-2022), fondateur du cadre théorique de l'A.S.I.

Dans quelle mesure la réussite à un problème de type A entraîne-t-elle la réussite à un problème de type B ?

sujets	A	B
S1	1	1
S2	0	0
S3	1	1
S4	1	1
S5	1	1
S6	0	1
S7	0	0
S8	0	1
S9	1	1
S10	1	1

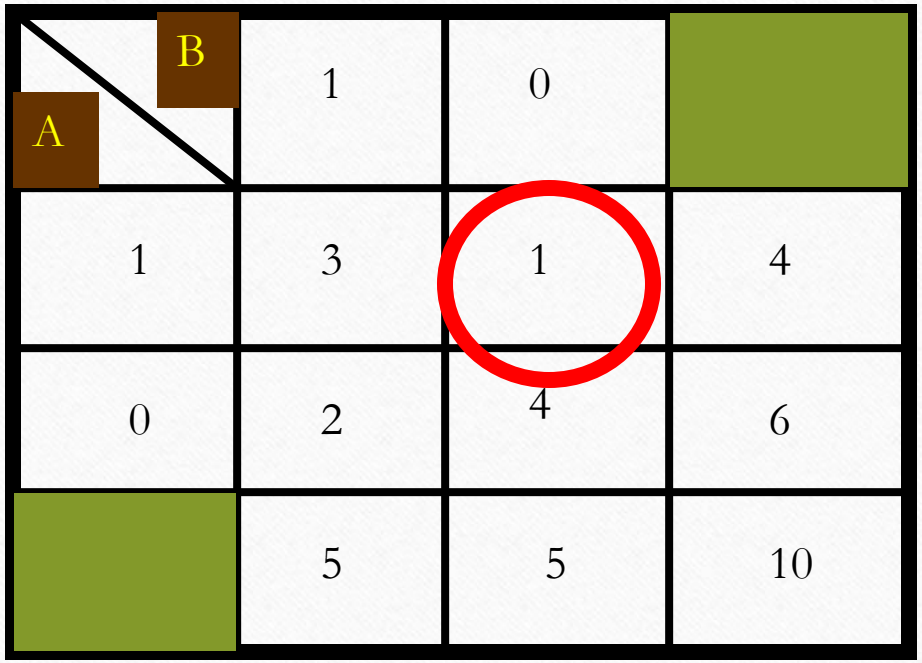


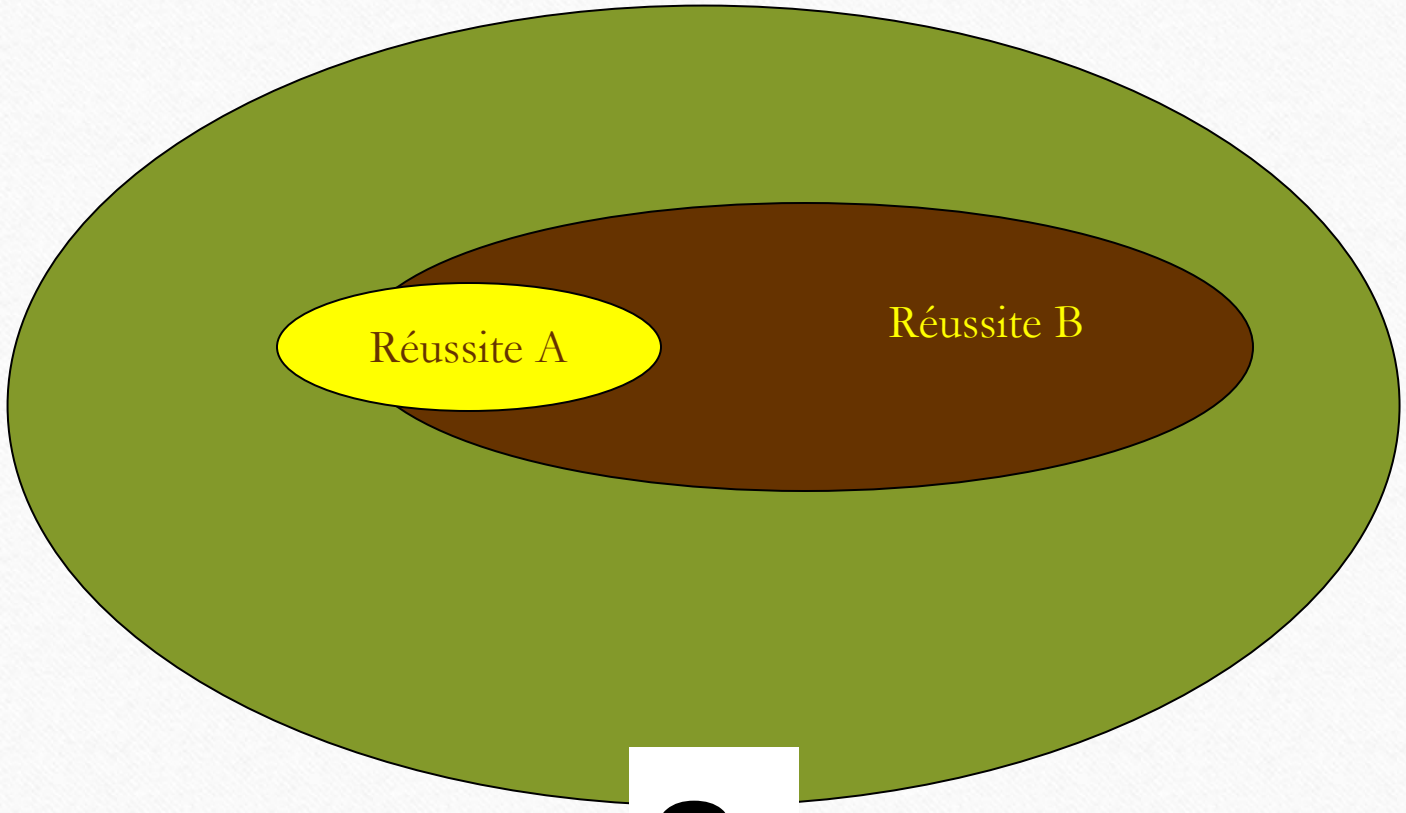
A \ B	1	0	
1	6	0	6
0	2	2	4
	8	2	10



[Réussite A] ⟶ [Réussite B]

sujets	A	B
S1	0	0
S2	0	1
S3	1	1
S4	1	0
S5	0	0
S6	1	1
S7	1	1
S8	0	0
S9	0	1
S10	0	0





[Réussite A] ? [Réussite B]

Enquête par Questionnaire

- Questionnaire et format des questions :

- Traitements initiaux : saisie dans un tableau de données

1 Vous et votre parcours antérieur de formation

[V1] Homme Femme (*entourez la réponse*)	**[V2]** Date de naissance :	*Codage pour anonymat :*

[V3] Vous êtes-vous déjà trouvé(e) confronté(e) à des situations qui demandent des connaissances en statistique ? :*(Entourer votre réponse)*

1_Jamais	2_Rarement	3_Quelquefois	4_Souvent	5_Très souvent	6_Presque chaque jour

[V4] Quand des connaissances en statistique sont requises, cela vous pose-t-il des problèmes ? : *(Entourer votre réponse)*

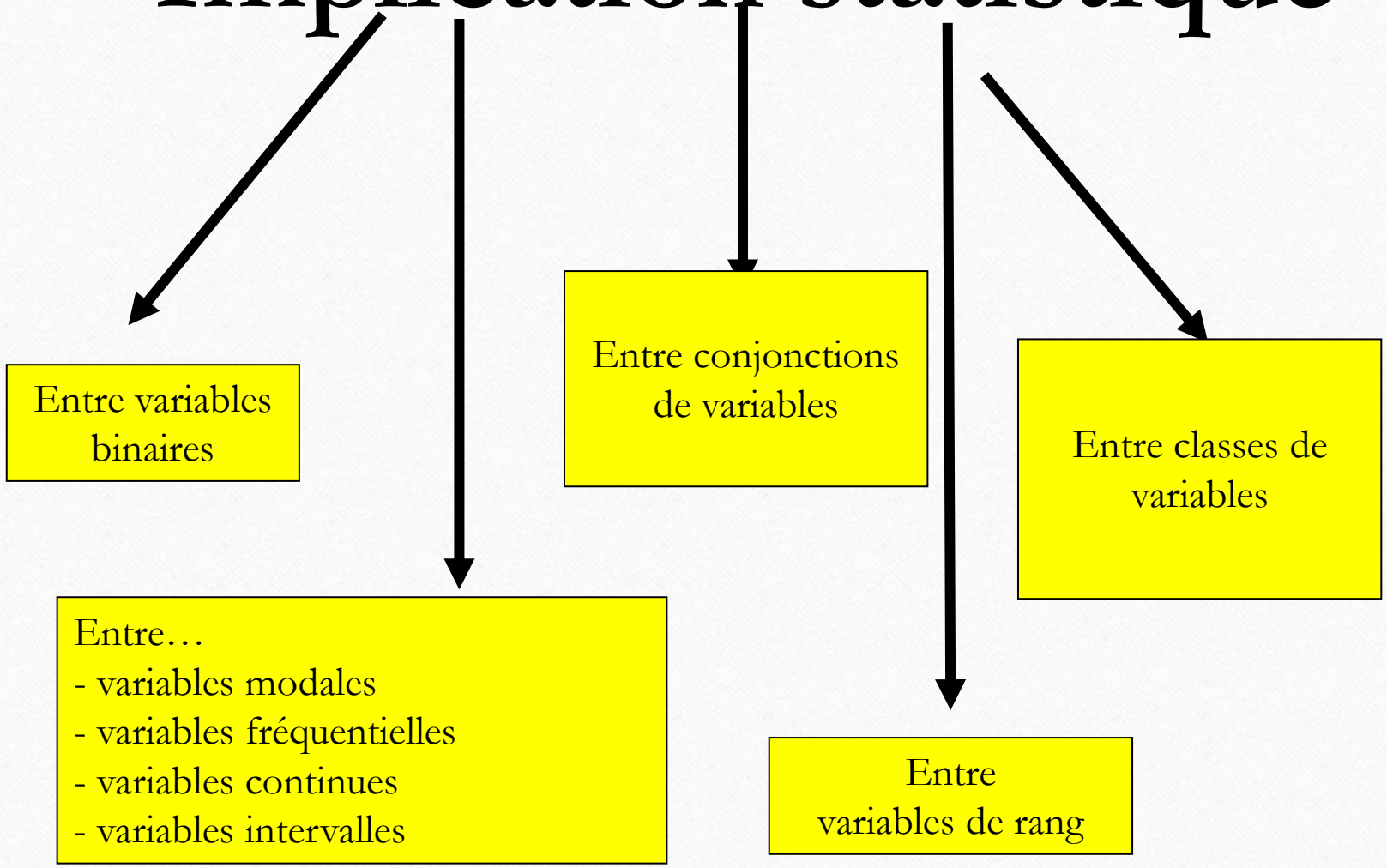
1_Pas du tout	2_Rarement	3_Quelquefois	4_Souvent	5_Très souvent	6_Toujours

Variable binaire : 0 ; 1

Age à la date de l'enquête

Variable modale : 0 ; 0,2 ; 0,4 ; 0,6 ; 0,8 ; 1

Implication statistique



Implication statistique

http://www.numdam.org/article/RSA_2005__53_1_5_0.pdf

Variables de rang

Tableau de contingence

A \ B	B est réalisé 1	B n'est pas réalisé 0	effectifs
A est réalisé 1	$n_{(A \text{ et } B)}$	$n_{(A \text{ et nonB})}$	$n_{(A)}$
A n'est pas réalisé 0	$n_{(\text{nonA et } B)}$	$n_{(\text{nonA et nonB})}$	$n_{(\text{nonA})}$
effectifs	$n_{(B)}$	$n_{(\text{nonB})}$	n

Effectif attendu sous l'hypothèse d'indépendance :

$$\frac{n_{(A)}\, n_{(\text{nonB})}}{n}$$

Tableau de contingence

A \ B	B est réalisé 1	B n'est pas réalisé 0	effectifs
A est réalisé 1	$n_{(A \text{ et } B)}$	$n_{(A \text{ et nonB})}$	$n_{(A)}$
A n'est pas réalisé 0	$n_{(\text{nonA et } B)}$	$n_{(\text{nonA et nonB})}$	$n_{(\text{nonA})}$
effectifs	$n_{(B)}$	$n_{(\text{nonB})}$	n

Effectif observé : $n_{(A \text{ et nonB})} < \frac{n_{(A)}\, n_{(\text{nonB})}}{N}$

Répulsion de A et nonB

Attraction de A et B

Tableau de contingence

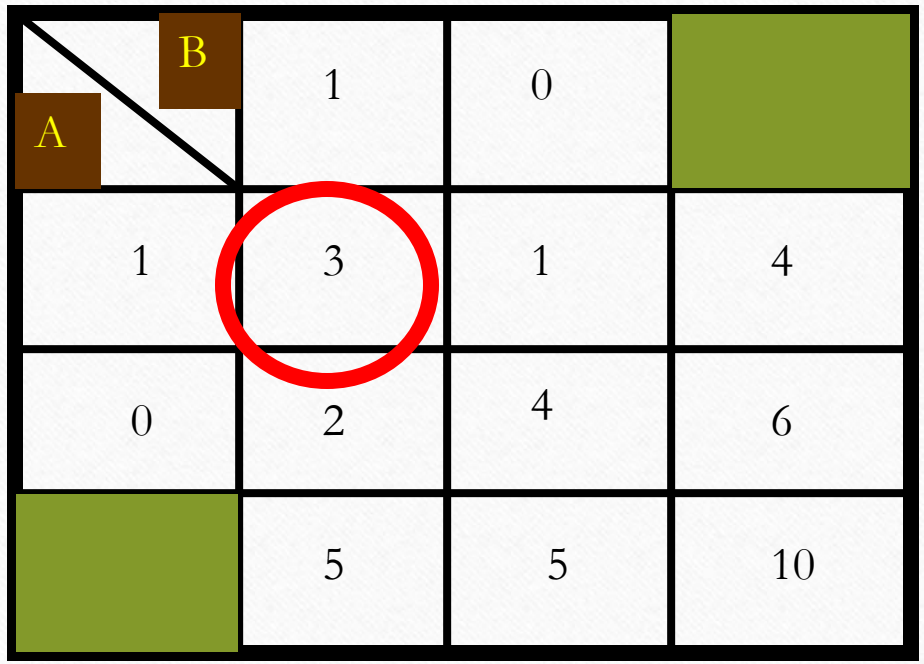
A \ B	B est réalisé 1	B n'est pas réalisé 0	effectifs
A est réalisé 1	$n_{(A \text{ et } B)}$	$n_{(A \text{ et nonB})}$	$n_{(A)}$
A n'est pas réalisé 0	$n_{(\text{nonA et } B)}$	$n_{(\text{nonA et nonB})}$	$n_{(\text{nonA})}$
effectifs	$n_{(B)}$	$n_{(\text{nonB})}$	n

Effectif observé : $n_{(A \text{ et nonB})} > \frac{n_{(A)}\ n_{(\text{nonB})}}{n}$

Répulsion de A et B

Attraction de A et nonB

sujets	A	B
S1	0	0
S2	0	1
S3	1	1
S4	1	0
S5	0	0
S6	1	1
S7	1	1
S8	0	0
S9	0	1
S10	0	0



Étude des Similarités

Tableau de contingence

A \ B	B est réalisé 1	B n'est pas réalisé 0	effectifs
A est réalisé 1	$n_{(A \text{ et } B)}$ similarité A et B	$n_{(A \text{ et nonB})}$ implication A vers B	$n_{(A)}$
A n'est pas réalisé 0	$n_{(\text{nonA et } B)}$ implication nonB vers nonA	$n_{(\text{nonA et nonB})}$ similarité nonA et nonB	$n_{(\text{nonA})}$
effectifs	$n_{(B)}$	$n_{(\text{nonB})}$	n

$$0 \leq n_{(A \text{ et nonB})} \leq n_{(A)} \leq n_{(B)} \leq n$$

$$0 \leq n_{(\text{nonA et } B)} \leq n_{(\text{nonB})} \leq n_{(\text{nonA})} \leq n$$

Ressources…

Références bibliographiques sur l'Analyse Statistique Implicative

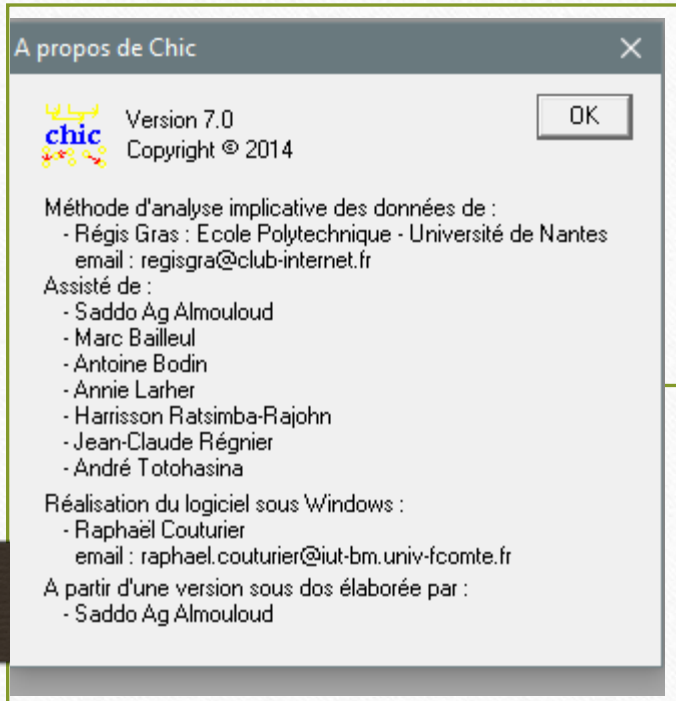
https://sites.univ-lyon2.fr/asi/ref/refasi

12 Colloques internationaux entre 2000 et 2023

https://sites.univ-lyon2.fr/asi/12/

Outils…

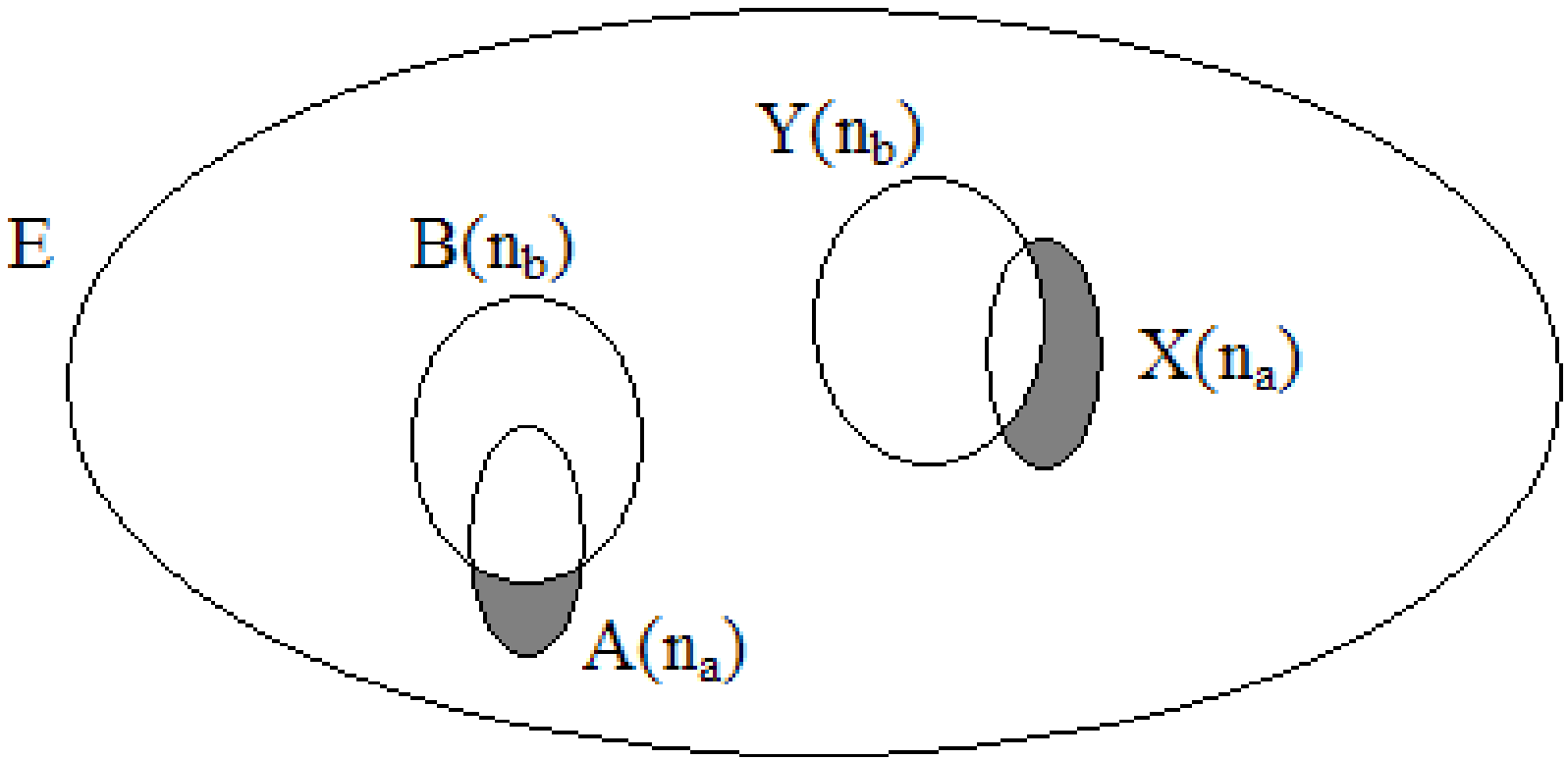
Outil logiciel



L ’implication statistique (2)

Quelques concepts fondateurs de la méthode d’analyse statistique implicative

Formalisation de la quasi-règle implicative dans l'approche classique



Les parties grisées représentent les contre-exemples à l'implication a⇒b

Fig. 1- *représentation par les diagrammes d'Euler*

		Variable b		
		1	0	**Total**
Variable a	1	$n(a \wedge b)$	$n(a \wedge \overline{b})$	$n(a)$
	0	$n(\overline{a} \wedge b)$	$n(\overline{a} \wedge \overline{b})$	$n(\overline{a})$
	Total	$n(b)$	$n(\overline{b})$	n

TAB 4 - *Tableau de contingence avec les notations ASI*

Définition 1: la quasi-règle $a \Rightarrow b$ est *admissible au niveau de confiance* $1-\alpha$ si et seulement si $\Pr\left[Card(X \cap \overline{Y}) \leq card(A \cap \overline{B})\right] \leq \alpha$

Intuitivement et qualitativement, ceci signifie que la quasi-implication $a \Rightarrow b$ sera admissible à l'issue d'une expérience si le nombre d'individus de E la contredisant est invraisemblablement petit par rapport au nombre d'individus attendu sous une hypothèse d'absence de lien.

Intensité d'implication

$\varphi(a;b)$

Définition 2: On appelle intensité d'implication de la quasi-règle $a \Rightarrow b$, le nombre $\varphi(a,b) = 1 - \Pr\left[Card(X \cap \overline{Y}) \leq Card(A \cap \overline{B})\right]$ si $n_b \neq n$

et

$\varphi(a.b) = 0$ si $n_b = n$

Indice d'implication

Définition 3: On appelle **indice d'implication** de la quasi-règle $a \Rightarrow b$, la variable aléatoire, notée $Q(a,\overline{b})$, déduite de la variable aléatoire $Card(X \cap \overline{Y})$ par centrage réduction.

$$Q(a,\overline{b}) = \frac{Card(X \cap \overline{Y}) - E[Card(X \cap \overline{Y})]}{\sigma[Card(X \cap \overline{Y})]}$$

L'intensité d'implication est une **valeur probabiliste**, et non une fréquence, qui fonde la décision de retenir ou non une relation de quasi-implication entre les variables binaires a et b.

Cette modélisation de la quasi-implication est pertinente pour **mesurer l'étonnement face au constat de la petitesse du nombre des contre-exemples en regard du nombre surprenant des instances de l'implication**. Il s'agit d'une mesure de la qualité inductive et informative de l'implication. Par conséquent, si la règle est triviale, comme dans le cas où B est très grand ou coïncide avec E, cet étonnement devient petit. Nous démontrons (Gras R., 1996 c) d'ailleurs que cette trivialité se traduit par une intensité d'implication très faible, voire nulle : *Si, n_a étant fixé et A étant inclus dans B, n_b tend vers n (B "croît" vers E), alors $\varphi(a,b)$ tend vers 0.*

C'est pourquoi nous définissons par « continuité »: $\varphi(a,b) = 0$ si $n_b = n$. De même, si $A \subset B$, $\varphi(a,b)$ peut être inférieure à 1 dans le cas où la confiance inductive, mesurée par l'étonnement statistique, est insuffisante.

Différents modèles pour évaluer l'intensité d'implication.

La détermination de **l'intensité d'implication** dépend du modèle retenu pour définir la loi de probabilité de la variable aléatoire « nombre de contre-exemples ».

Modèle binomial (1)

Examiner la qualité de la quasi-règle a⇒b, dans le cas où les variables sont binaires, revient à mesurer de façon équivalente celle de l'inclusion du sous-ensemble des transactions satisfaisant a dans le sous-ensemble des transactions satisfaisant b. Les contre-exemples relatifs à l'inclusion sont en effet les mêmes que ceux qui sont relatifs à l'implication exprimée par : « toute transaction satisfaisant a satisfait aussi b ». Dans cette optique ensembliste, dès que $n_a \leq n_b$, la qualité de la quasi-règle a⇒b, ne peut qu'être sémantiquement meilleure que celle de $b \Rightarrow a$. Nous supposerons donc, par la suite, que $n_a \leq n_b$ lors de l'étude de a⇒b. Dans ce cas, la population-mère est finie et $Card\, E = n$.

Modèle binomial (2)

La modélisation binomiale fut chronologiquement la première adoptée. Rappelons brièvement en quoi consiste **le modèle binomial**. Avec les notations adoptées, X et Y sont deux sous-ensembles aléatoires, indépendamment choisis dans l'ensemble des parties de E, respectivement de mêmes cardinaux n_a et n_b que les sous-ensembles de réalisations de a et de b. La valeur observée $n_{a \wedge \bar{b}}$ peut être considérée comme la réalisation d'une variable aléatoire $\mathrm{Card}\,(X \cap \overline{Y})$ qui représente le nombre aléatoire de contre-exemples à l'inclusion de X dans Y, contre-exemples observés au cours de n tirages successifs indépendants.

Modèle binomial (3)

De là, Card $(X \cap \overline{Y})$ peut être considérée comme une variable binomiale de paramètres n et π où π est elle-même estimée par $p = \frac{n_a}{n}\frac{n_{\overline{b}}}{n}$.

Ainsi :

$$\Pr\left[Card(X \cap \overline{Y}) = k\right] = C_n^k \left(\frac{n_a n_{\overline{b}}}{n^2}\right)^k \left(1 - \frac{n_a n_{\overline{b}}}{n^2}\right)^{n-k}$$

Modèle binomial (4)

La· variable· centrée· réduite· estimée· $Q(a,\bar{b})$ · admet· alors· comme· réalisation°:

$$q(a,\bar{b}) = \frac{n_{a\wedge\bar{b}} - \frac{n_a\, n_{\bar{b}}}{n}}{\sqrt{\frac{n_a\, n_{\bar{b}}}{n}\left(1 - \frac{n_a\, n_{\bar{b}}}{n^2}\right)}}$$

Modèle binomial (5)

Nous obtenons l'intensité d'implication empirique estimée :

$$\varphi(a,b) = 1 - \Pr\left[Q(a,\bar{b}) \le q(a,\bar{b})\right] = 1 - \sum_{0}^{n_{a\wedge\bar{b}}} C_n^k \left(\frac{n_a n_{\bar{b}}}{n^2}\right)^k \left(1 - \frac{n_a n_{\bar{b}}}{n^2}\right)^{n-k}$$

NB : La loi de probabilité de $Q(a,\bar{b})$ peut être approchée par celle de la loi de Laplace-Gauss centrée réduite N(0,1).

Remarque : Nous pouvons noter que l'indice d'implication est nul si et seulement les deux variables a et b sont indépendantes. En effet

$$q(a,\bar{b}) = \frac{n_{a\wedge\bar{b}} - \frac{n_a\, n_{\bar{b}}}{n}}{\sqrt{\frac{n_a\, n_{\bar{b}}}{n}\left(1 - \frac{n_a\, n_{\bar{b}}}{n^2}\right)}} = 0 \Leftrightarrow n_{a\wedge\bar{b}} - \frac{n_a\, n_{\bar{b}}}{n} = 0$$

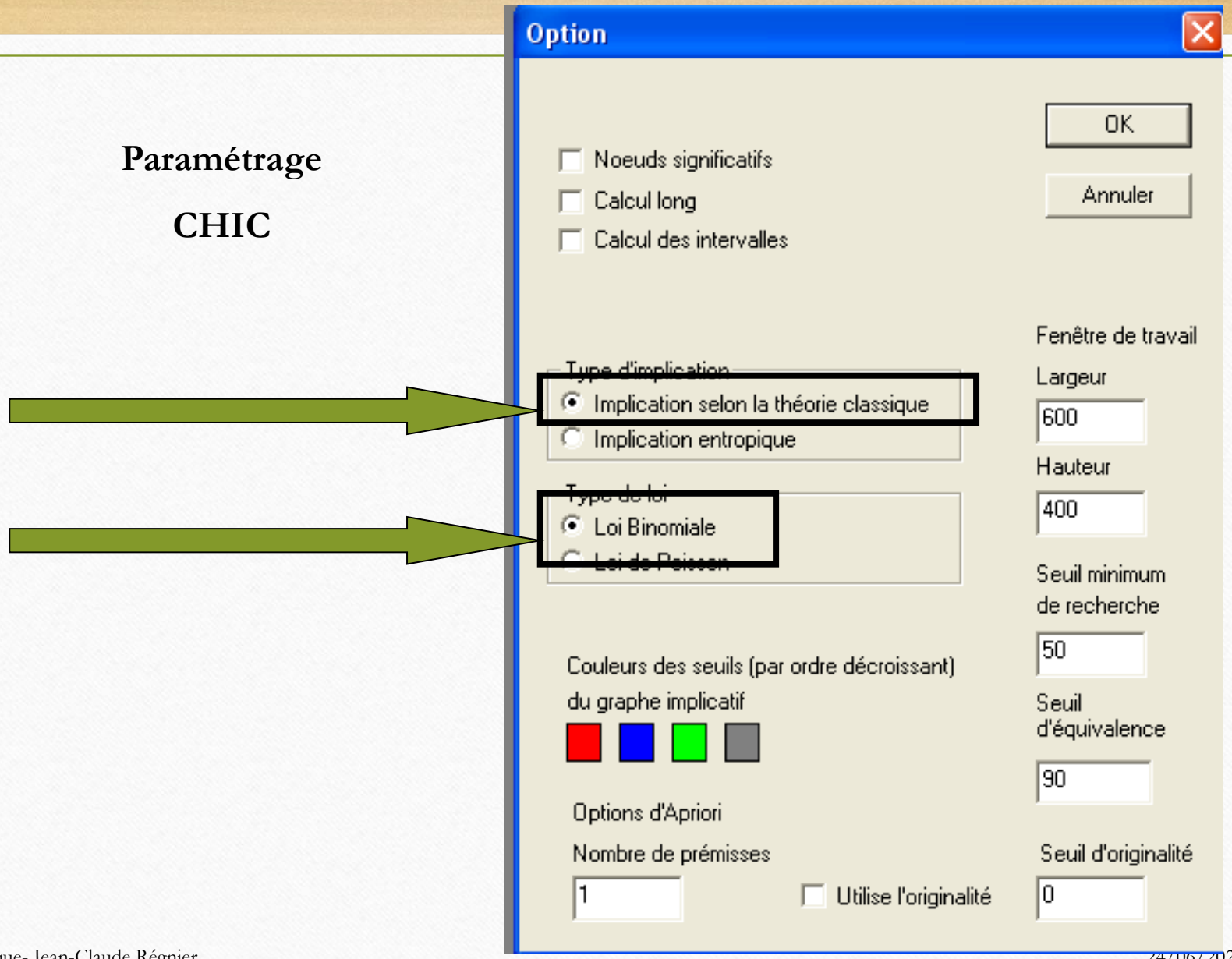
$$q(a,\bar{b}) = 0 \Leftrightarrow n_{a\wedge\bar{b}} = \frac{n_a\, n_{\bar{b}}}{n}$$ ou encore

$$q(a,\bar{b}) = 0 \Leftrightarrow \frac{n_{a\wedge\bar{b}}}{n} = \frac{n_a}{n}\frac{n_{\bar{b}}}{n}$$

Cette dernière relation traduit la propriété d'indépendance statistique.

Paramétrage

CHIC



Modèle de Poisson (1)

Dans ce modèle, nous considérons que la variable « nombre de contre-exemples » $Card(X \cap \overline{Y})$ suit la distribution de probabilités de la variable de Poisson de paramètre λ

$$\forall s \in \{0,1,2,\ldots,n\} \; Pr[Card(X \cap \overline{Y}) = s] = \frac{\lambda^s}{s!} e^{-\lambda}$$

Dans notre situation λ est remplacé par son estimation :

$$\hat{\lambda} = \lambda_{estimé} = \frac{n_a . n_{\overline{b}}}{n}$$

Modèle de Poisson (2)

En conséquence, la probabilité pour que le hasard conduise, sous l'hypothèse d'absence de lien a priori entre a et b, à au plus le nombre de contre-exemples observé, est estimée par :

$$\Pr\left[Card(X \cap \overline{Y}) \leq n_{a \wedge \overline{b}}\right] = \sum_{s=0}^{n_{a \wedge \overline{b}}} \frac{\hat{\lambda}^s}{s!} . e^{-\hat{\lambda}}$$

L'intensité de l'implication estimée, notée $\varphi(a,b)$ est alors :

$$\varphi(a,b) = 1 - \Pr\left[Card(X \cap \overline{Y}) \leq n_{a \wedge \overline{b}}\right] = 1 - \sum_{s=0}^{n_{a \wedge \overline{b}}} \frac{\hat{\lambda}^s}{s!} . e^{-\hat{\lambda}}$$

Modèle de Poisson (3)

La **variable** « indice d'implication » est dans ce cas

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

La **valeur empirique de l'indice :**

$$q(a,\bar{b}) = \frac{n_{a \wedge \bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$$

Lien entre l'indice d'implication dans le modèle de Poisson et la mesure du χ^2

Proposition 3 Considérons la mesure χ^2 et l'indice d'implication $q(a,\bar{b})$ entre les variables binaires a et b alors $\frac{\chi^2}{q(a,\bar{b})^2} = \frac{n^2}{n_a n_{\bar{b}}}$

puisque

$$q(a,\bar{b}) = \frac{n_{a\wedge\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} = \frac{n_a n_b - n n_{a\wedge b}}{\sqrt{n n_a n_{\bar{b}}}}$$

et

$$\chi^2 = \frac{n(n_{a\wedge b} n_{\bar{a}\wedge\bar{b}} - n_{\bar{a}\wedge b} n_{a\wedge\bar{b}})^2}{n_a n_b n_{\bar{a}} n_{\bar{b}}}$$

Modèle de Poisson (4)

$$Q(a,\bar{b}) = \frac{Card(X \cap \bar{Y}) - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}}$$ qui résulte du centrage-réduction de la variable de Poisson, $Card(X \cap \overline{Y})$, peut être approchée par la variable gaussienne centrée réduite $N(0;1)$.

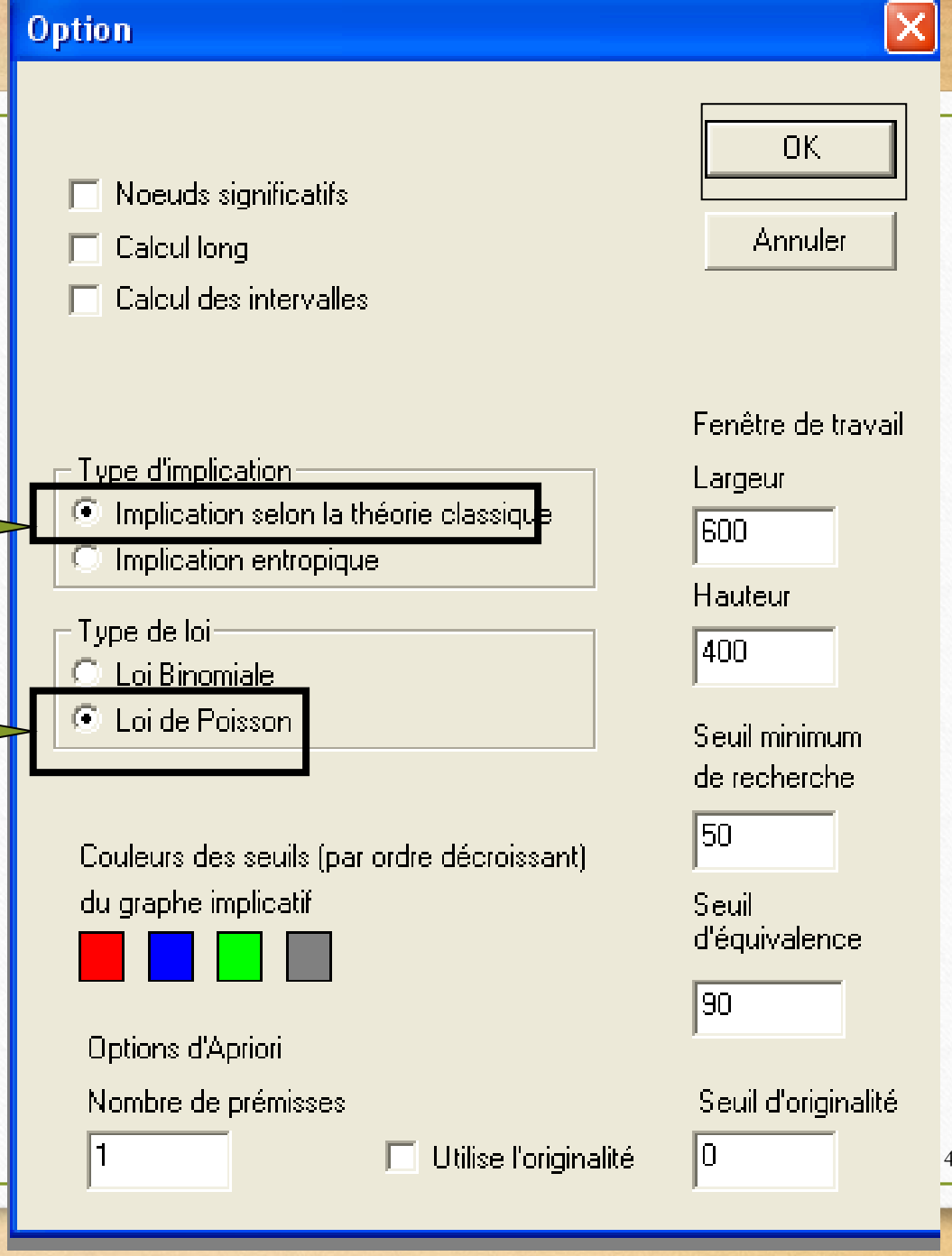
Modèle de Poisson (5)

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

Insistons sur le sens de cette intégrale. Elle représente la probabilité gaussienne pour que le nombre de transactions observées satisfaisant la quasi-règle $a \Rightarrow b$, soit supérieur à celui qui serait observable sous l'hypothèse d'indépendance de a et b. Autrement dit, $\Pr[Q(a,\bar{b}) \leq q(a,\bar{b})]$ est la **p-value** du test visant à réfuter l'hypothèse de l'indépendance de a et b au profit d'une relation de type quasi-implication

Paramétrage

CHIC

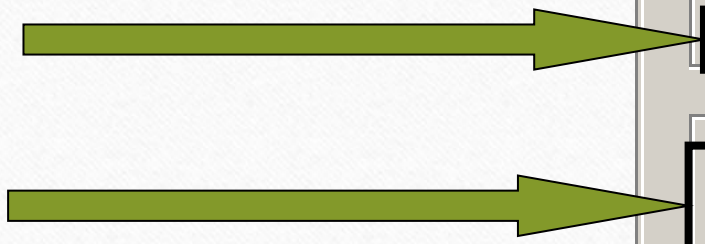


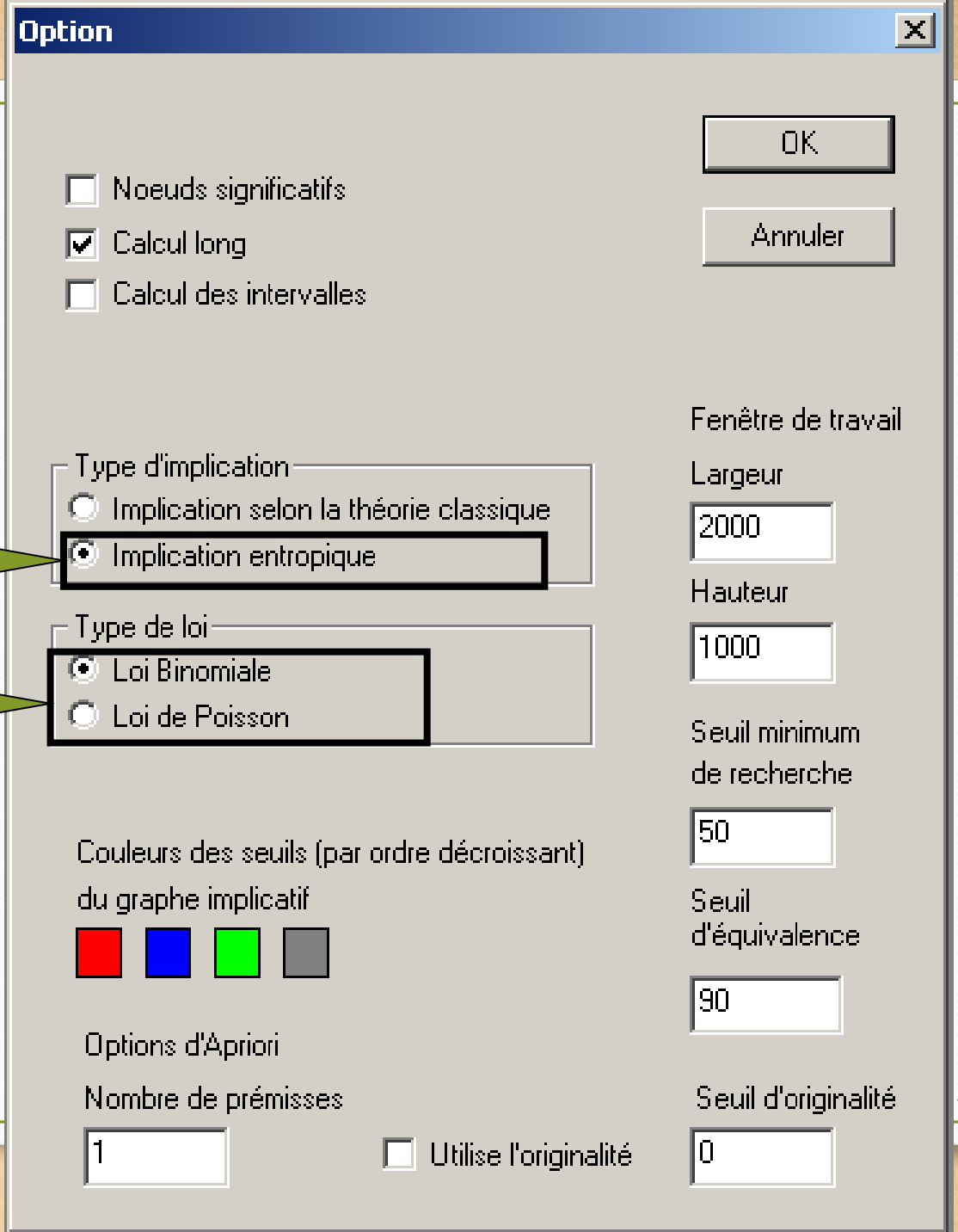
L ’implication statistique (3)

Quelques concepts fondateurs de la méthode d’analyse statistique implicative

Paramétrage

CHIC





Formalisation de la quasi-règle implicative dans l'approche entropique (1)

Nous utilisons à la fois l'intensité d'implication et un autre indice qui rend compte de la dissymétrie entre les situations S1=(a et b) et S'1=(a et non b) qui concerne la quasi-règle a⇒b ainsi que celle entre les situations S2=(non a et non b) et S'2 =(a et non b) qui concerne la quasi-règle contraposée.

Ce sont les mêmes instances qui contredisent la quasi-implication et sa contraposée

Formalisation de la quasi-règle implicative dans l'approche entropique (2)

Fondée sur le concept d'entropie de Shannon (1949), l'entropie conditionnelle relative à S1 et S'1 quand a est réalisée :

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

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

Formalisation de la quasi-règle implicative dans l'approche entropique (3)

Posons $\alpha = \frac{n_a}{n}$ la fréquence de a, $\overline{\beta} = \frac{n_{\overline{b}}}{n}$ la fréquence de non b et $t = \frac{n_{a \wedge \overline{b}}}{n}$ la fréquence des contre-exemples. Nous construisons alors deux fonctions h_1 et h_2 définies dans [0 ;1] comme suit :

$$h_1(t) = H(b/a) = -(1-\frac{t}{\alpha})\log_2(1-\frac{t}{\alpha}) - \frac{t}{\alpha}\log_2\frac{t}{\alpha} \text{ si } t\in[0,\frac{\alpha}{2}[\text{ et } h_1(t) = 1 \text{ si } t\in[\frac{\alpha}{2}, \alpha]$$

$$h_2(t) = H(\overline{a}/\overline{b}) = -(1-\frac{t}{\overline{\beta}})\log_2(1-\frac{t}{\overline{\beta}}) - \frac{t}{\overline{\beta}}\log_2\frac{t}{\overline{\beta}} \text{ si } t\in[0,\frac{\overline{\beta}}{2}[\text{ et } h_2(t) = 1 \text{ si } t\in[\frac{\overline{\beta}}{2}, \overline{\beta}]$$

Formalisation de la quasi-règle implicative dans l'approche entropique (4)

Définition 1: L' **indice d'inclusion de A**, support de a, **dans B**, support de b, est le nombre qui intègre l'information délivrée par la réalisation du nombre de contre-exemples, d'une part à la quasi-règle a⇒b et, d'autre part, à la quasi-règle non(b)⇒non(a)

$$i(a,b) = [(1 - h_1^2(t))(1 - h_2^2(t))]^{\frac{1}{4}}$$

Formalisation de la quasi-règle implicative dans l'approche entropique (5)

Définition 2 : **L'intensité d'implication-inclusion** ou **intensité entropique**, est le nombre suivant:

$$\psi(a,b) = [i(a,b)\ \varphi(a,b)]^{1/2}$$

Formalisation de la quasi-règle implicative dans l'approche implifiance [implication+ confiance] (1)

Définition 3 : On appelle **implifiance** la mesure de l'implication statistique qui prend en compte l'implication directe et sa contraposée ainsi que la confiance en chacune de ces deux formes inclusives.

$$\Phi(a,b) = \varphi(a,b)[C_1(a,b)C_2(\bar{b},\bar{a})]^{1/4}$$

avec la confiance

$$C_1(a,b) = Fr[Y|X] = \frac{Card(X \cap Y)}{Card(X)} = \frac{n_{a\cap b}}{n_a}$$

$$C_2(\bar{b},\bar{a}) = Fr[\bar{X}|\bar{Y}] = \frac{Card(\bar{X} \cap \bar{Y})}{Card(\bar{Y})} = \frac{n_{\bar{a}\cap\bar{b}}}{n_{\bar{b}}}$$

Extension de l'Analyse Statistique Implicative aux variables non binaires

relation de propension entre deux variables modales

Considérons pour cela, n couples de variables aléatoires X_i et Y_i indépendantes et identiquement distribuées, pour tout sujet i, dont la moyenne et la variance estimées sont celles des observations empiriques de E. Les réalisations de ces variables sont respectivement les n couples (a_i, b_i) de valeurs de $[0, 1]^2$.

$$m_a = \frac{1}{n}\sum_{i \in E} a_i \qquad \nu_a = \frac{1}{n}\sum_{i \in E} (a_i - m_a)^2$$

L'indice empirique d'implication devient :

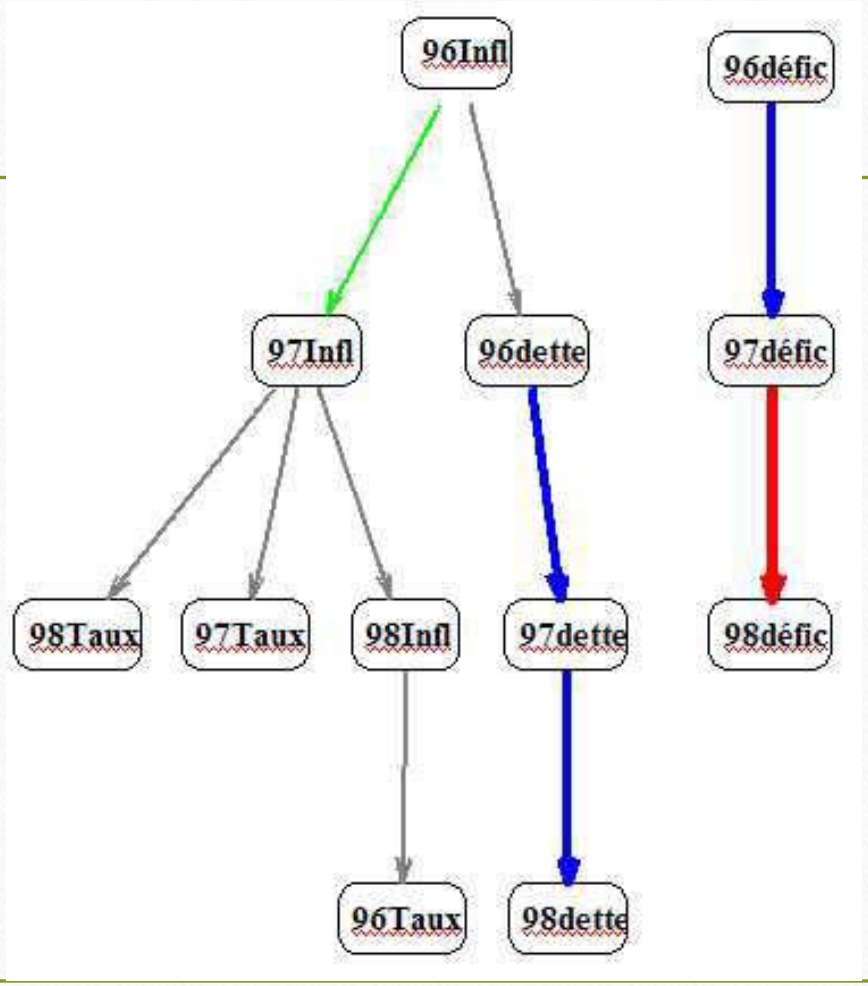
$$\tilde{q}(a,\bar{b}) = \frac{\frac{1}{n}\sum_{i \in E} a_i \bar{b}\, i - m_a m_{\bar{b}}}{\sqrt{\frac{\nu_a(\nu_b + m_{\bar{b}}^2) + m_a^2 \nu_b}{n}}}$$

Quant à l'estimation de l'intensité de propension, elle est encore obtenue par :

$$\varphi(a,b) = 1 - \Pr\left[\tilde{Q}(a,\bar{b}) \le \tilde{q}(a,\bar{b})\right] = \frac{1}{\sqrt{2\pi}} \int_{\tilde{q}(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}}\, dt$$

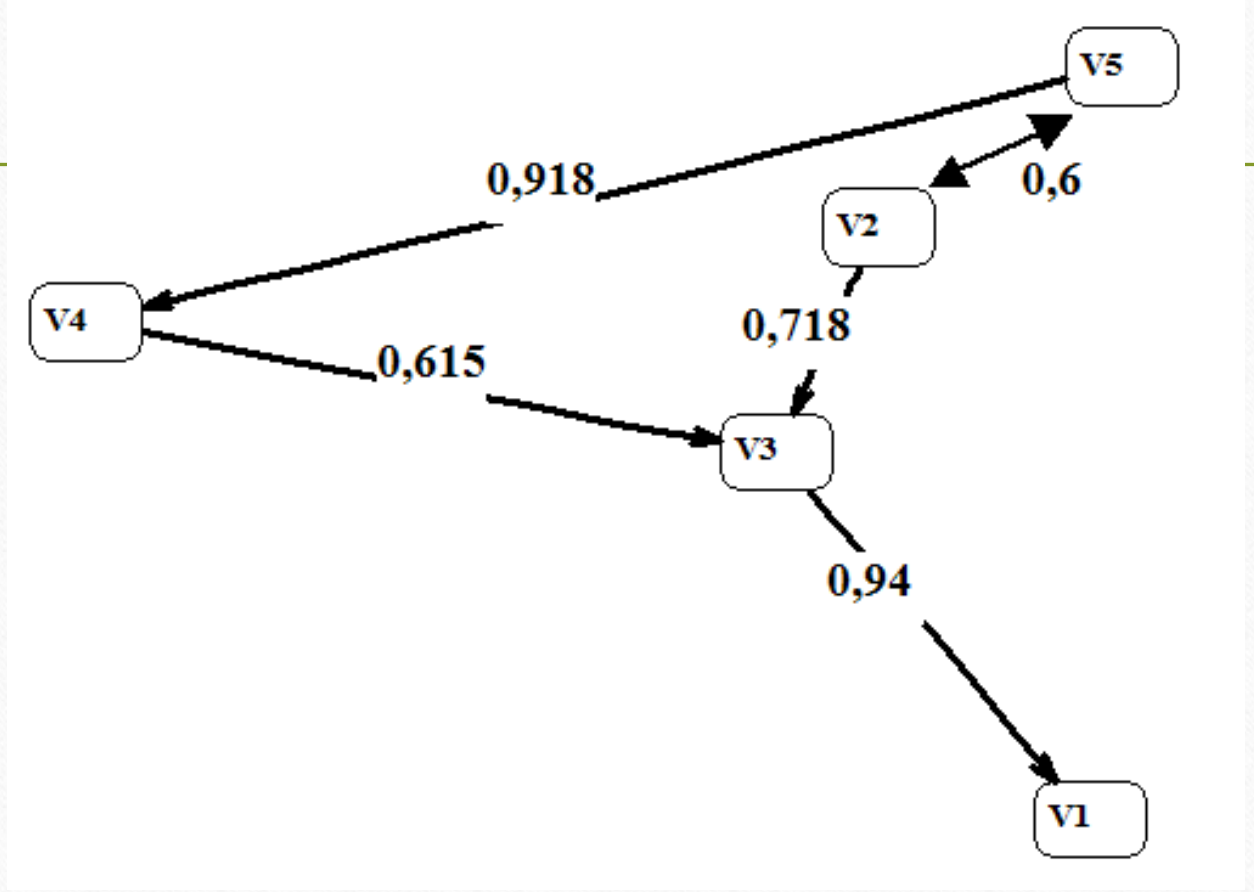
Représentations graphiques

Graphe implicatif



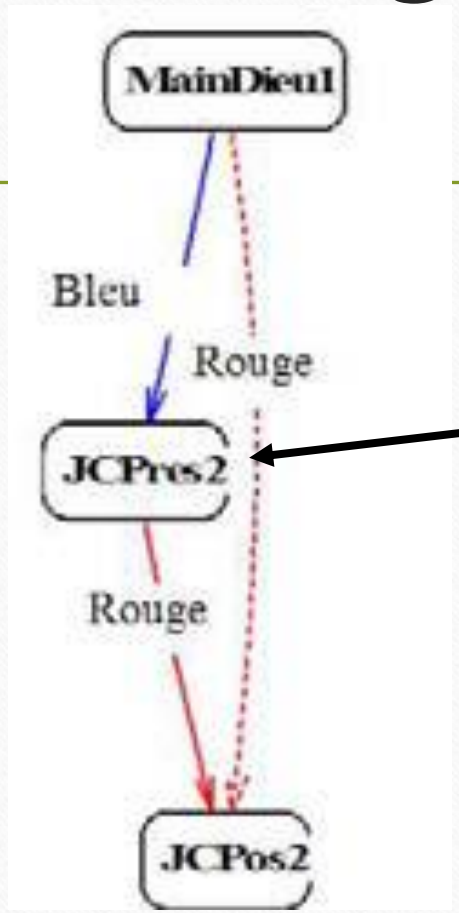
Représentations graphiques

Graphe implicatif



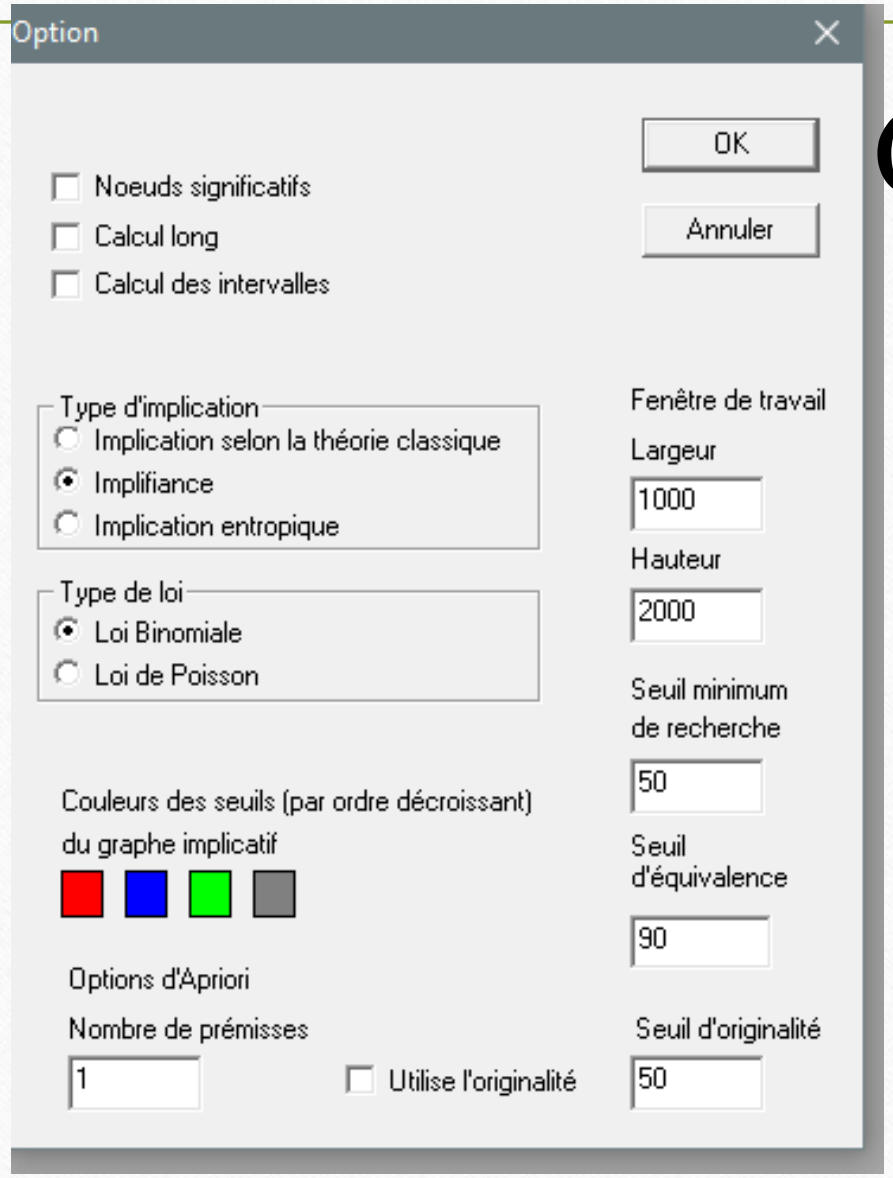
Représentations graphiques

Graphe implicatif



Fermeture transitive

$$\varphi(a,b) > 0.50$$

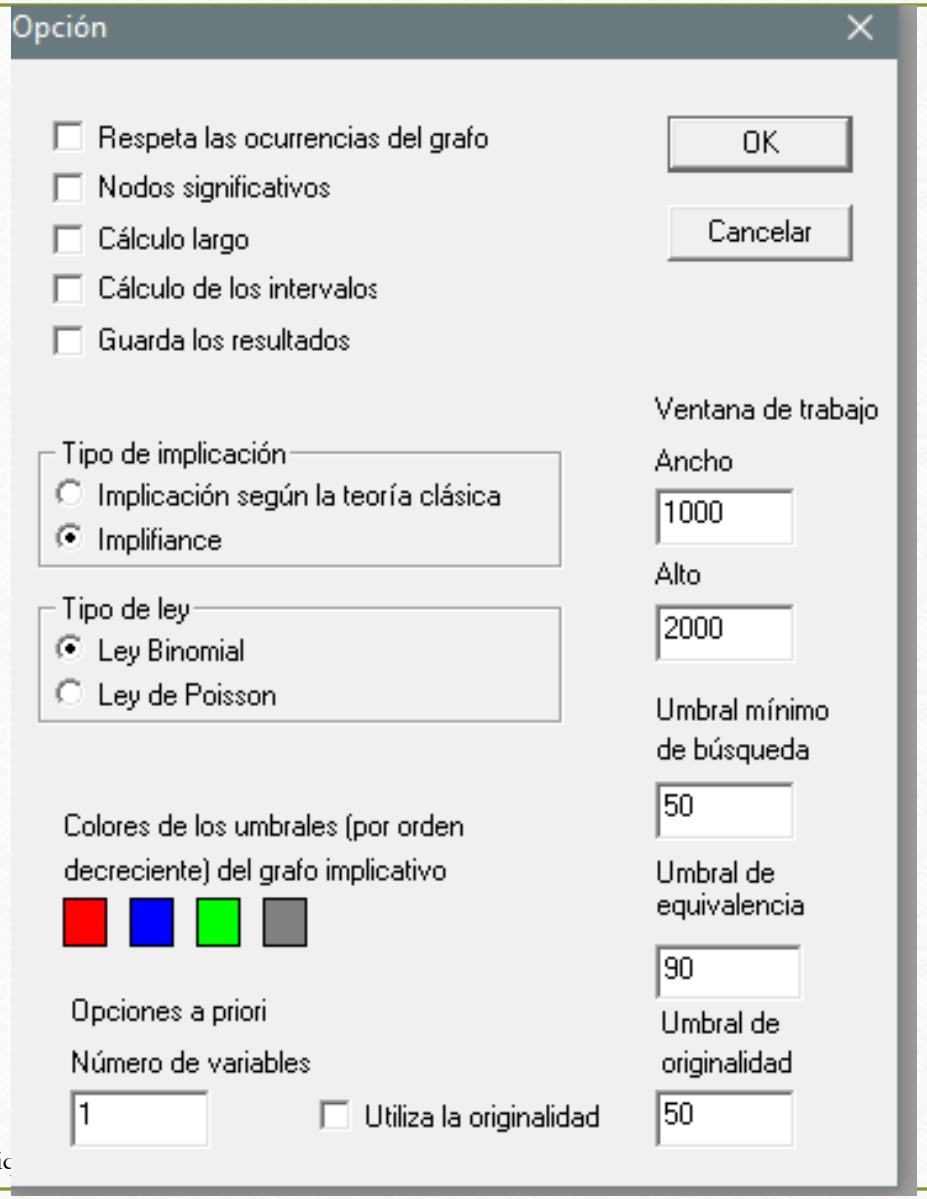


C.H.I.C. 7

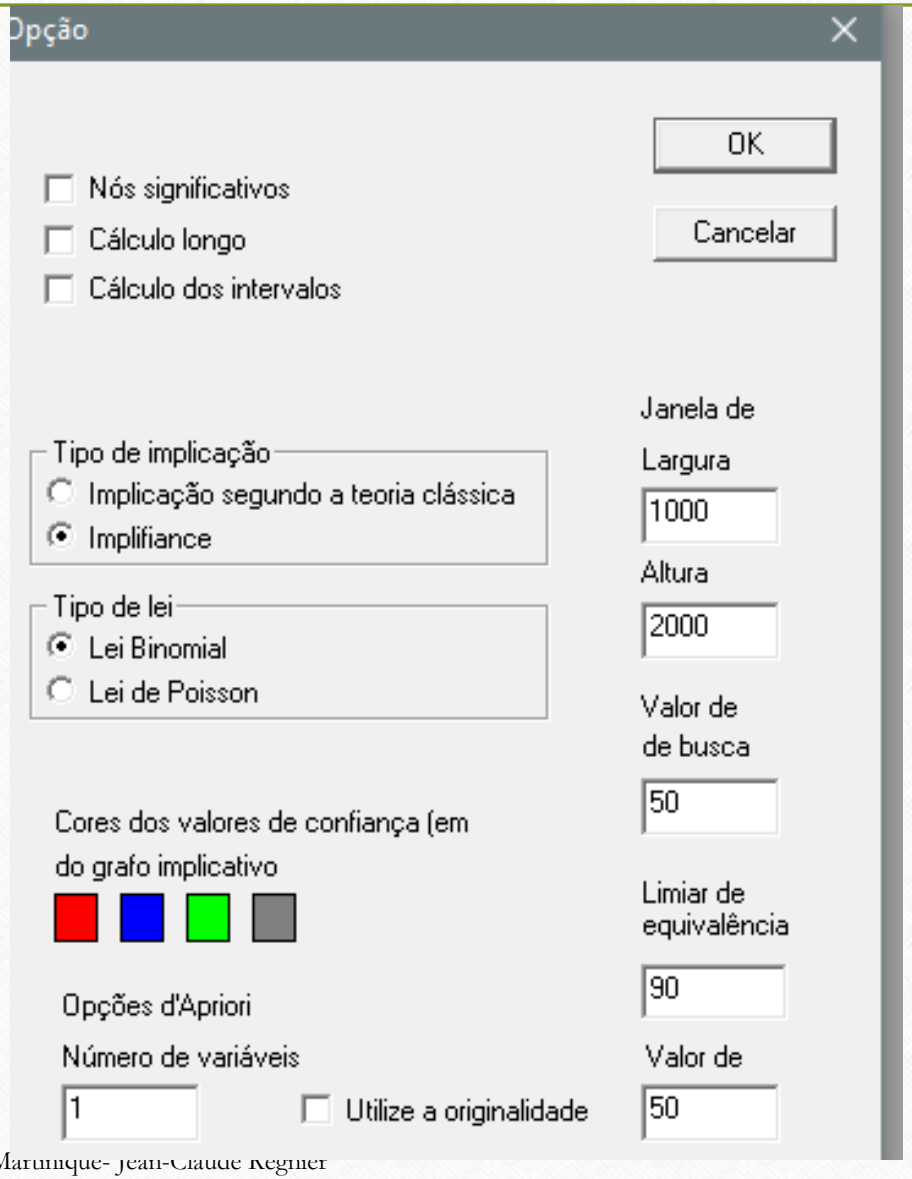
Idiome français

C.H.I.C. 7

Idiome espagnol



C.H.I.C. 7

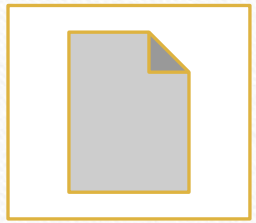


Idiome portugais

Thèse de doctorat en Sciences de l'éducation et de la formation (2023)

Mohamed Boumoudjou

Exemple



Base de données

Enseignement-apprentissage des mathématiques par la résolution de problèmes à l'école primaire en contexte plurilingue au Maroc : prise en compte des ressources mathématiques mobilisées dans des pratiques quotidiennes.

Approches pédagogique, didactique et ethnomathématique.

Classification hiérarchique cohésive ou cohésitive orientée

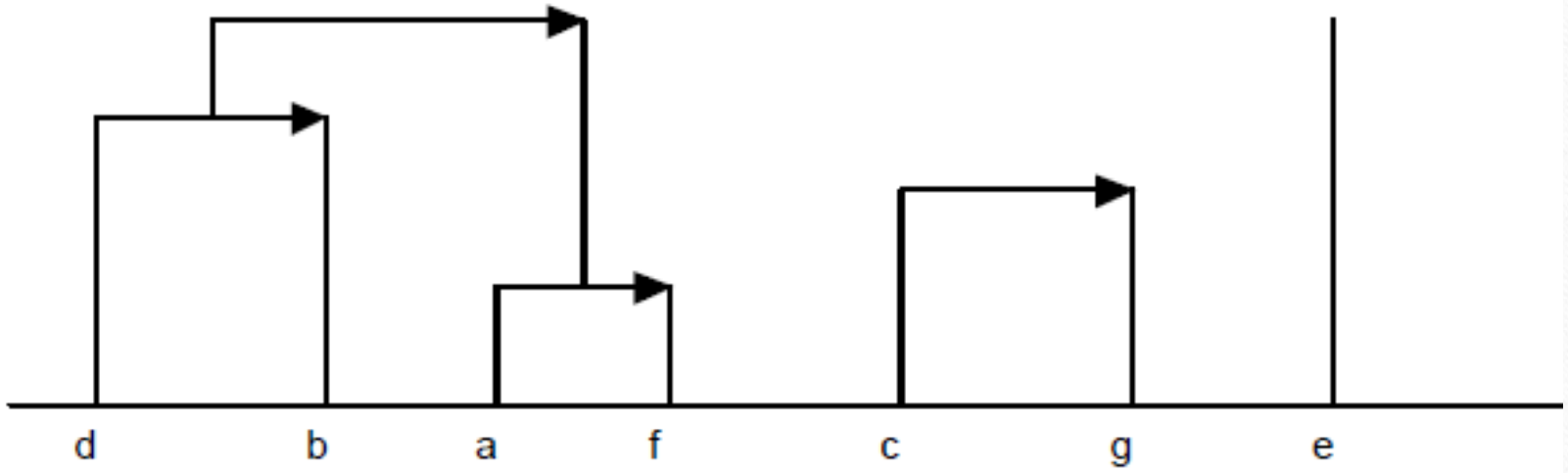


FIG. 10 *Un exemple de hiérarchie orientée*

Indice de cohésion degré 1 et 2

Définition de l'indice de cohésion de degré 1 d'une classe à un élément{a}

$$coh(a) = 1$$

Définition de l'indice de cohésion de degré 2 d'une classe à deux éléments {a,b}

$$coh(a,b) = (1 - \left[-p(\log_2(p)) - (1-p)(\log_2(1-p))\right]^2)^{\frac{1}{2}} \text{ où } p = \varphi(a,b) \geq 0.50$$

$$coh(a,b) = 0 \; si \; p = \varphi(a,b) < 0.5$$

Indice de cohésion degré r

Définition de l'indice de cohésion de degré r d'une classe C à r éléments $\{a_1, a_2, \ldots, a_r\}$

$$coh(C) = \left[\prod_{i=1,\ldots,r-1}^{j=2,\ldots,r\ j>i} coh(a_i, a_j) \right]^{\frac{2}{r(r-1)}}$$

Indice de cohérence d'une classe

Définition: La ***cohérence*** *o(C)* d'une classe *C* d'une hiérarchie orientée est définie par la probabilité $Pr(I > i)$.

Variable I = nombre d'inversions dans une permutation. Comparer l'ordre des éléments de la classe orientée par la relation de quasi-implication et l'ordre des effectifs dans la construction du graphe implicatif

Indice *CO* de cohésion-cohérence d'une classe C_k

Définition : L'indice *CO* de cohésion-cohérence qui mesure la significativité de la classe C_{k+1} formée au niveau k + 1 est défini par :

$$co(C_{k+1}) = \frac{c(C_{k+1})}{c(C_k)} . o(C_{k+1})$$