

De la qualité des règles d'association:

Étude comparative des mesures M_{GK} et *Confiance*.

Totohasina André, Feno Daniel Rajaonasy

Département de mathématiques et informatique. ENSET.
Université d'Antsiranana
Antsiranana
Madagasikara
{totohasina, fenodaniel2}@yahoo.fr

RÉSUMÉ. Le présent papier donne des propriétés mathématiques et statistiques intéressantes et éclairantes de la mesure M_{GK} de Guillaume-Kentchaff. Parallèlement à cela, il permet de comparer M_{GK} et la mesure pionnière *Confiance* en fouille de données quant à la pertinence des règles produites. Un exemple d'application sur des données concrètes sera exposé à l'appui.

ABSTRACT. The present paper talks about interesting and highlighting mathematical and statistical properties of the so-called Guillaume-Kenchaff quality measure M_{GK} . It allows comparison between M_{GK} and the traditional measure *Confiance* about pertinence of produced rules. We shall talk about an application on a real data.

MOTS-CLÉS : Mesure de qualité, M_{GK} , *Confiance*, Pertinence, Règle d'association positive ou négative, Implication statistique, Situations de référence, Application.

KEYWORDS : Quality measure, M_{GK} , Pertinence, Association rule, Statistical implication, Reference situations, Application

1. Motivations, travaux existants sur M_{GK} et modélisation

Remarquons tout d'abord que cette mesure a été proposée indépendamment dans (1) en 2000 par inspiration à la mesure de Loevinger et dans (2) en 2004. Signalons au passage que, grâce à ses propriétés mathématiques, cette mesure a déjà reçu diverses dénominations différentes selon les auteurs, à savoir *ION* par (3) démontrant sa propriété d'implication orientée normalisée, *CPIR* (conditional probability increment ratio) par Wu et al (2004) (2) par sa formule qui exprime un taux d'incrément de probabilité conditionnelle et par son efficacité dans l'extraction des règles d'association non redondantes, enfin *Conf_G* (confiance de Guillaume) par S. Ferré (cf. p.139-140 de (4) montrant qu'elle est moins ambiguë et plus intelligible donc plus appropriée que la traditionnelle *Confiance* d'Agrawal et al.(5)(qui ne distingue pas attraction et répulsion) dans la recherche d'implications approximatives pour constituer un système d'information logique, CF (Certainty Factor, facteur de certitude). En fait, très peu de travaux parlent de cet intéressant indice de qualité. Aussi, s'avère-t-il crucial de l'approfondir.

Par rapport à plusieurs modélisations probabilistes utilisées dans la littérature (voir par exemple (6)), on se place ici dans le cadre d'un contexte de la fouille de données binaires $\mathbb{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$, où $\mathcal{O}$ est un ensemble fini d'entités ou d'objets, $\mathcal{A}$ un ensemble fini non vide d'attributs (ou de variables, ou items, ou des propriétés) et $\mathcal{R}$ une relation binaire de $\mathcal{O}$ vers $\mathcal{A}$. L'appartenance d'un couple $(o, a) \in \mathcal{O} \times \mathcal{A}$ au graphe de la relation $\mathcal{R}$ exprime le fait que l'objet o possède la propriété a , tout attribut pouvant être identifié à une application de $\mathcal{O}$ dans $\{0, 1\}$, où la valeur 1 mesure la présence de l'attribut chez un objet de $\mathcal{O}$. Notons n la cardinalité de $\mathcal{O}$ ($n = |\mathcal{O}|$). Tout sous-ensemble X de $\mathcal{A}$ s'appelle un *motif* ou *itemset* de $\mathcal{A}$, et sa négation logique notée $\bar{X}$ le *motif négatif associé au motif* X , et tout élément de $\mathcal{O}$ un *objet* ou une *entité* de $\mathcal{O}$. Pour tout motif X de $\mathcal{A}$, remarquons les huit points suivants : $\forall a \in \mathcal{A}, \bar{a}$, i.e. $(1 - a)$ identifie l'absence de l'attribut a chez une entité ; $\forall e \in \mathcal{O}, X(e) = 1 \iff \forall a \in X, e\mathcal{R}a$, i.e. $a(e) = 1$; soit $X = \bigwedge_{a \in X} a =$ la conjonction des présences d'un nombre fini d'attributs de $\mathcal{A}$; $\bar{X} = \bigvee_{a \in X} (1 - a) =$ la disjonction des absences des attributs de X ; $X' = \{e \in \mathcal{O} | \forall x \in X, e\mathcal{R}x\}$, i.e. le dual ou l'*extension* de X de $\mathcal{A}$; de façon duale, pour une partie $E \subseteq \mathcal{O}$, l'*intension* de E notée E' définie par $E' = \{a \in \mathcal{A} | \forall e \in E, a(e) = 1\}$, c'est l'ensemble des attributs communs aux objets de E ; $\bar{X}' = \mathcal{O} - X' = \bar{X}'$; cette coïncidence renforce l'appellation de motif négatif pour $\bar{X}$; $\bar{\bar{X}} = X$: on retrouve ainsi le caractère involutif de l'action de négation en logique formelle, i.e. la loi de De Morgan ; $X \subseteq \mathcal{A}$, mais $\bar{X} \not\subseteq \mathcal{A}$.

Il est facile de voir que pour deux motifs X et Y du contexte, on a : $(X \cap Y)' = X' \cap Y'$ et $(X \cup Y)' = X' \cup Y'$. Or a priori on n'a pas de raison pour le réfuter, nous nous plaçons dans le cadre d'une hypothèse d'équiprobabilité des événements élémentaires de $\mathcal{O}$. D'où la considération de l'espace probabilisé discret $(\mathcal{O}, \mathcal{P}(\mathcal{O}), P)$, P étant la probabilité uniforme. Par conséquent, pour tout X de $\mathcal{A}$, en notant $n_X = |X'|$ la cardinalité de X' , $Supp(X) = \frac{n_X}{n}$ représente une estimation de la probabilité $P(X')$ de l'événement X' . De plus, par glissement de terminologie qui se justifie par la dualité entre exten-

sion et intension, il s'avère naturel d'adopter les définitions suivantes : Deux motifs sont indépendants (resp. dépendants), si leurs extensions respectifs sont indépendants (resp. dépendants) dans l'espace probabilisé $(\mathcal{O}, \mathcal{P}(\mathcal{O}), P)$. La littérature atteste que ceci a déjà été longtemps intuitivement et implicitement accepté.

2. Propriétés de la mesure M_{GK} : construction, seuils de signification, relation avec coefficient de corrélation linéaire.

Selon la modélisation probabiliste adoptée dans notre approche, on peut facilement retrouver cette mesure par une démarche plus directe.

Remarque 1 On obtient facilement les doubles inégalités suivantes, pour deux motifs deux motifs X et Y :

- (i) Si X favorise Y , (i.e. $P(Y|X') > P(Y')$), alors $0 < P(Y|X') - P(Y') \leq 1 - P(Y')$.
- (ii) Si X défavorise Y , (i.e. $P(Y|X') < P(Y')$) alors $-P(Y') \leq P(Y|X') - P(Y') < 0$.
- (iii) " X défavorise Y " équivaut à " X favorise $\bar{Y}$ "; donc $1 - P(Y') < 1 - P(Y'|X')$ si et seulement si $P(\bar{Y}') < P(\bar{Y}'|X')$.

Il en résulte que d'une part : $-1 \leq \frac{P(Y'|X') - P(Y')}{P(Y')} \leq 0$, si X défavorise Y ; et d'autre part : $0 \leq \frac{P(Y'|X') - P(Y')}{1 - P(Y')} \leq 1$, si X favorise Y . D'où, la définition suivante.

Définition 1 Soit X et Y deux motifs d'un contexte de fouille de données. On définit :

$$M_{GK}(X \rightarrow Y) = \begin{cases} \frac{P(Y'|X') - P(Y')}{1 - P(Y')}, & \text{si } X \text{ favorise } Y \\ \frac{P(Y'|X') - P(Y')}{P(Y')}, & \text{si } X \text{ défavorise } Y. \end{cases} \quad (1)$$

Tout d'abord, notons que pour deux motifs X et Y non indépendants, des deux choses l'une : soit il y a attraction mutuelle, auquel cas la dépendance est positive, soit il y a répulsion, et alors il existe une dépendance positive entre les deux motifs X et $\bar{Y}$ et l'on considère $X \rightarrow \bar{Y}$ d'une part, puis entre $\bar{X}$ et Y et l'on considère $\bar{X} \rightarrow Y$ d'autre part. Dans les deux cas, on aura toujours à considérer une dépendance positive. La mesure M_{GK} se compose d'une composante favorable M_{GK}^f et d'une défavorable M_{GK}^d comme suit :

$$M_{GK}(X \rightarrow Y) = \begin{cases} M_{GK}^f(X \rightarrow Y), & \text{si } X \text{ favorise } Y \\ M_{GK}^d(X \rightarrow Y), & \text{si } X \text{ défavorise } Y. \end{cases}$$

Ainsi, c'est surtout la composante favorable M_{GK}^f qui va guider la sémantique de M_{GK} .

Définition 2 Une mesure de qualité μ est dite normalisée si elle vérifie les cinq conditions suivantes : pour toute règle d'association $X \rightarrow Y$, on a : (i) $\mu(X \rightarrow Y) = -1$,

si $P(Y'/X') = 0$; (ii) $-1 < \mu(X \rightarrow Y) < 0$, si $0 \neq P(Y'/X') < P(Y')$ (i.e. X et Y sont négativement dépendants (en répulsion partielle)); (iii) $\mu(X \rightarrow Y) = 0$, si $P(Y'/X') = P(Y')$ (i.e. X et Y sont indépendants); (iv) $0 < \mu(X \rightarrow Y) < 1$, si $1 \neq P(Y'/X') > P(Y')$, i.e. si X favorise Y ou X et Y s'attirent partiellement; (v) $\mu(X \rightarrow Y) = 1$, si $P(Y'/X') = 1$ (ou encore si X implique totalement Y).

Entre l'indice pionnier *Confiance* et M_{GK} , on a la proposition ci-dessous.

Proposition 1 (1) Contrairement à *Confiance*, M_{GK} est normalisée; Aux marginales fixées, *Confiance* et M_{GK}^f sont deux fonctions croissantes du nombre d'exemples de règle, cependant M_{GK}^f croît plus lentement que *Confiance*.

(2) $\forall (X \rightarrow Y)$ telle que X favorise Y , on a :

(a) Si leurs extensions sont telles que $X' \subseteq Y'$, alors $\text{conf}(X \rightarrow Y) = M_{GK}^f(X \rightarrow Y) = 1$: $X \rightarrow Y$ est dite une règle exacte.

(b) Si $X' \not\subseteq Y'$, alors $0 < M_{GK}^f(X \rightarrow Y) < \text{conf}(X \rightarrow Y) < 1$, ou encore $\frac{1 - M_{GK}^f}{1 - \text{conf}}(X \rightarrow Y) > 1$: $X \rightarrow Y$ est dite une règle approximative.

Corollaire 1 M_{GK} a favorablement un pouvoir discriminant plus élevé que *Confiance*.

De la Proposition 1 et du Corollaire 1, on tire le corollaire suivant.

Corollaire 2 Pour un même type de règle, la mesure de qualité normalisée M_{GK} produit moins de règles que *Confiance*.

Proposition 2 (i) Si X favorise Y , nous avons la relation d'équivalence des deux règles contraposées :

$$M_{GK}^f(\overline{Y} \rightarrow \overline{X}) = M_{GK}^f(X \rightarrow Y). \quad (2)$$

(ii) Si X défavorise Y , nous avons la relation :

$$M_{GK}^d(\overline{Y} \rightarrow \overline{X}) = \frac{P(X')P(Y')}{(1 - P(X'))(1 - P(Y'))} M_{GK}^d(X \rightarrow Y) \quad (3)$$

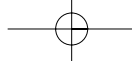
Ainsi, compte tenu de la remarque faite ci-dessus, on peut considérer que M_{GK} est favorablement implicative, alors que *Confiance* n'est pas implicative. La Proposition 3 ci-dessous montre que la mesure de qualité M_{GK} est favorablement non symétrique.

Proposition 3 (i) Si X favorise Y , nous avons la relation :

$$M_{GK}^f(Y \rightarrow X) = \frac{1 - P(Y')}{1 - P(X')} \frac{P(X')}{P(Y')} M_{GK}^f(X \rightarrow Y). \quad (4)$$

(ii) Si X défavorise Y , nous avons la relation :

$$M_{GK}^d(Y \rightarrow X) = M_{GK}^d(X \rightarrow Y) \quad (5)$$



Pour ce qui concerne les règles négatives à droite, nous avons :

Proposition 4 Pour deux motifs positifs X et Y , on a l'égalité et l'équivalence suivante :

$$M_{GK}^f(X \rightarrow \bar{Y}) = -M_{GK}^d(X \rightarrow Y). \quad (6)$$

$$\text{Et } \forall \alpha \in]0, 1[, \text{ on a : } (-1 < M_{GK}^d(X \rightarrow Y) < -\alpha \iff \alpha < M_{GK}^f(X \rightarrow \bar{Y}) < 1) \quad (7)$$

Ainsi, plus le degré de quasi-incompatibilité entre deux motifs est élevé, plus la qualité de la règle négative à droite correspondante est favorablement meilleure. Alors que l'équivalence permet d'élaguer directement toute candidate règle négative à droite dont la valeur M_{GK} de la règle positive associée est négative et située dans l'intervalle $[-\alpha, 0[$, pour un seuil α fixé dans $]0, 1[$. Au sujet des règles négatives à gauche, on a la

Proposition 5 Pour deux motifs positifs X et Y , on a les égalités suivantes.

(1) Si X défavorise Y (donc, X favorise $\bar{Y}$ et aussi $\bar{X}$ favorise Y), alors :

$$M_{GK}^f(\bar{X} \rightarrow Y) = \frac{P(X')}{1 - P(X')} \frac{P(Y')}{1 - P(Y')} M_{GK}^f(X \rightarrow \bar{Y}) \quad (8)$$

(2) Si X favorise Y (donc, X défavorise $\bar{Y}$ et aussi $\bar{X}$ défavorise Y), alors :

$$M_{GK}^d(\bar{X} \rightarrow Y) = \frac{P(X')}{(1 - P(X'))} \frac{1 - P(Y')}{P(Y')} M_{GK}^d(X \rightarrow \bar{Y}) \quad (9)$$

À partir des deux Propositions 4 et 5 précédentes, on tire les relations entre une règle négative à gauche et la règle positive correspondante.

Corollaire 3 Pour deux motifs positifs X et Y , on a les égalités suivantes.

(1) Si X défavorise Y (donc, X favorise $\bar{Y}$ et aussi $\bar{X}$ favorise Y), alors :

$$M_{GK}^f(\bar{X} \rightarrow Y) = -\frac{p(X')}{1 - P(X')} \frac{P(Y')}{1 - P(Y')} M_{GK}^d(X \rightarrow Y) \quad (10)$$

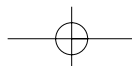
(2) Si X favorise Y (donc, X défavorise $\bar{Y}$ et aussi $\bar{X}$ défavorise Y), alors :

$$M_{GK}^d(\bar{X} \rightarrow Y) = -\frac{P(X')}{(1 - P(X'))} \frac{1 - P(Y')}{P(Y')} M_{GK}^f(X \rightarrow Y) \quad (11)$$

De ce corollaire, on tire aisément une relation (cf. Corol. 4) entre seuils pour sélectionner des règles négatives à gauche intéressantes ou élaguer celles non intéressantes.

Corollaire 4 Pour deux motifs positifs X et Y , on a les égalités suivantes. Si X défavorise Y (donc, X favorise $\bar{Y}$ et aussi $\bar{X}$ favorise Y), alors pour un seuil $\alpha \in]0, 1[$ fixé :

$$-1 < M_{GK}^d(X \rightarrow Y) < -\alpha \iff \frac{p(X')}{1 - P(X')} \frac{P(Y')}{1 - P(Y')} \alpha < M_{GK}^f(\bar{X} \rightarrow Y) < \frac{p(X')}{1 - P(X')} \frac{P(Y')}{1 - P(Y')}$$



Proposition 6 Aux motifs emboîtés, on a les relations ci-dessous :

- (i) Si $X \subseteq Y \subseteq Z \subseteq \mathcal{A}$ alors $M_{GK}(X \rightarrow Z) = M_{GK}(X \rightarrow Y)M_{GK}(Y \rightarrow Z)$.
(ii) Si $X_1 \subseteq X_2 \subseteq \dots \subseteq X_i \subseteq X_{i+1} \subseteq \dots \subseteq X_p \subseteq \mathcal{A}$ alors $M_{GK}(X_1 \rightarrow X_p) = \prod_{i=1}^{p-1} M_{GK}(X_i \rightarrow X_{i+1})$

Par conséquent, M_{GK} est multiplicative sur une chaîne du treillis de motifs. Ce qui apparaît intéressant dans la pratique, sur le plan d'élagage grâce à la transitivité axiale.

Corollaire 5 Soient $X_1, X_2, \dots, X_i, X_{i+1}, \dots, X_p$ des motifs tels que $X_1 \subseteq X_2 \subseteq \dots \subseteq X_i \subseteq X_{i+1} \subseteq \dots \subseteq X_p$.

- (i) Si $X_1 \rightarrow X_p$ est (M_{GK}, α) -valide alors $\forall i, j \in \{1, \dots, p\}$ avec $i < j$, $X_i \rightarrow X_j$ est (M_{GK}, α) -valide.
(ii) Si il existe $i, j \in \{1, \dots, p\}$ tels que $X_i \rightarrow X_j$ est non (M_{GK}, α) -valide alors $\forall l, k \in \{1, \dots, p\}$ tels que $l \leq i$ et $j \leq k$, $X_l \rightarrow X_k$ est aussi non (M_{GK}, α) -valide.

A part les cinq situations de référence mentionnées ci-dessus, Blanchard et al. (7) considèrent une autre situation de référence à savoir la situation d'équilibre ou d'incertitude maximale (i.e., $|X' \cap Y'| = |X' \cap \bar{Y}'|$) : Une mesure de qualité est dite "mesure de déviation d'équilibre" si elle prend une valeur constante quand le nombre d'exemples et de contre-exemples de la règle sont égaux (7). Or à l'équilibre, $M_{GK}^f(X \rightarrow Y) \approx \frac{1}{2}$ asymptotiquement (3). Ainsi, M_{GK} est une mesure de déviation d'équilibre pour un grand volume de données. Ce qui n'est pas le cas pour *Confiance*. Par ailleurs, par la considération du tableau de contingence obtenu par le croisement des deux motifs, s'appuyant sur la statistique de χ^2 de Pearson à un degré de liberté, la Proposition 7 ci-dessous montre qu'il est facile d'élaborer les abaques des valeurs critiques pour la signification d'une règle d'association M_{GK} -valide. X et Y étant deux motifs du contexte, on a :

Proposition 7

$$M_{GK}(X \rightarrow Y) = \begin{cases} \sqrt{\frac{1}{n} \frac{n-n_X}{n_X} \frac{n_Y}{n-n_Y} \chi^2}, & \text{si } X \text{ fav. } Y \text{ ou } X \text{ et } Y \text{ sont indépds} \\ -\sqrt{\frac{1}{n} \frac{n-n_X}{n_X} \frac{n_Y}{n-n_Y} \chi^2}, & \text{si } X \text{ défav. } Y \text{ ou } X \text{ et } Y \text{ sont indépds} \end{cases}$$

La figure 1 plus bas, par exemple, fournit plus d'éclairage sur les modes d'évaluation du degré de dépendance entre deux motifs, dans cinq situations de référence : dépendance positive (Figure 1 (1)), dépendance négative (figure 1 (2)), indépendance (figure 1 (3)), incompatibilité (figure 1 (4)), et l'implication logique (figure 1 (5)). Il apparaît que, contrairement à χ^2 , la mesure M_{GK} calcule le degré de dépendance orientée sur une échelle limitée dans l'intervalle $[-1, +1]$, en plus de son orientation : par exemple, dans le tableau (figure 1 (2)) la très significative dépendance entre les deux motifs révélée par χ^2 est en fait une dépendance négative et comme $M_{GK}^f(X \rightarrow \bar{Y}) = 0, 1$, donc très faible devant $M_{GK}^f(\bar{X} \rightarrow Y) = 4/9 = 0, 444, 1$, seule la règle négative à gauche $\bar{X} \rightarrow Y$ est significativement valide. Notons également qu'ici la $\text{conf}(\bar{X} \rightarrow Y) = 0, 75$ est suffisamment

élevée aussi, alors que son évaluation selon M_{GK} reste relativement faible (donc à taux de saturation(M_{GK}) = $\frac{1-M_{GK}}{1-0} = \frac{5}{9} \simeq 0.556 > (\frac{1-conf}{1-0} = \frac{1}{4} = 0.25)$). En fait, il est immédiat qu'en cas de lien positif non totalement implicatif, $0 < \frac{M_{GK}^f}{conf} < 1$. Ce qui conduit bien à affirmer que M_{GK} est plus discriminante que la confiance. Considérons deux motifs X et Y du contexte $\mathbb{K}$. Leurs *extensions* respectives X' et Y' sont deux événements de $\mathcal{P}(\mathcal{O})$, tels que : l'espérance mathématique de $\mathbf{1}_{X'}$ étant $E(\mathbf{1}_{X'}) = P(X')$ et sa variance $\sigma_{\mathbf{1}_{X'}}^2 = P(X')P(\overline{X'})$, la covariance de X et Y est $cov(X, Y) = cov(\mathbf{1}_{X'}, \mathbf{1}_{Y'}) = E(\mathbf{1}_{X'} \times \mathbf{1}_{Y'}) - E(\mathbf{1}_{X'})E(\mathbf{1}_{Y'}) = P(X' \cap Y') - P(X')P(Y')$. D'où la relation clé :

$$M_{GK}(X \rightarrow Y) = \begin{cases} \rho(X, Y) \sqrt{\frac{n_X}{n_X} \frac{n_Y}{n_Y}}, & \text{si } X \text{ favorise } Y \text{ ou } X \text{ et } Y \text{ sont indépendants} \\ \rho(X, Y) \sqrt{\frac{n_X}{n_X} \frac{n_Y}{n_Y}}, & \text{si } X \text{ défavorise } Y \text{ ou } X \text{ et } Y \text{ indépendants} \end{cases}$$

Et réciproquement :

$$\rho(X, Y) = \sqrt{\frac{P(X')P(Y')}{P(X')P(Y')}} M_{GK}^f(X \rightarrow Y) \mathbf{1}_f(X \rightarrow Y) + \sqrt{\frac{P(X')P(Y')}{P(X')P(Y')}} M_{GK}^d(X \rightarrow Y) \mathbf{1}_d(X \rightarrow Y).$$

Ainsi, la mesure M_{GK} et le coefficient de corrélation linéaire ρ ont le même signe. Aussi, M_{GK} intègre la corrélation et oriente cette dépendance eu égard à sa propriété non symétrique. D'où les diverses manières d'interpréter une règle M_{GK} -valide intégrant celle du coefficient de corrélation linéaire $\rho(X, Y)$ à la différence de la non symétrie cette fois : en terme de similarité dynamique, une dynamique vers l'identité de la variable prémisses sur la variable conséquent, en terme de moyenne géométrique.

3. Conclusion

Ces propriétés montrent que la mesure de qualité normalisée M_{GK} permet de sélectionner moins de règles que la mesure Confiance si l'on se borne aux règles positives, mesure conjointement l'écart à l'indépendance et le degré d'implication statistique entre deux motifs, respecte la condition d'équilibre pour un grand volume de données. S'avérant favorablement plus discriminante que *Confiance*, M_{GK} a entre autres l'avantage d'élaguer systématiquement les règles dont prémisses et conséquent sont indépendants. Grâce à sa cohérence avec l'attraction et la répulsion entre deux motifs, elle est moins ambiguë et plus intelligible que le test de Khi-deux d'indépendance et la traditionnelle *Confiance*. Néanmoins, eu égard au caractère intuitif du mot *confiance* dans le langage social et au concept de probabilité conditionnelle qu'on lui fait incarner, il s'avère utile de garder ce terme, mais cette fois pour les règles M_{GK} -valides uniquement. Dans le futur proche, un travail sur la génération des bases composites de telles règles d'association sera effectué. D'ors et déjà, en guise de validation, nous exhiberons une application sous forme de bases composites de règles sur un jeu de données concrètes (Cf. Annexe).

$X \setminus Y$	Y	$\bar{Y}$	
X	3000	2000	5000
$\bar{X}$	2500	2500	5000
	5500	4500	10000

(1) Dépendance positive
 $\chi^2 = 101$, $M_{GK} = +0.11$

$X \setminus Y$	Y	$\bar{Y}$	
X	1000	3000	4000
$\bar{X}$	4500	1500	6000
	5500	4500	10000

(2) Dépendance négative
 $\chi^2 = 2424$, $M_{GK} = -0.54$

$X \setminus Y$	Y	$\bar{Y}$	
X	2200	1800	4000
$\bar{X}$	3300	2700	6000
	5500	4500	10000

(3) Indépendance
 $\chi^2 = 0$, $M_{GK} = 0$

$X \setminus Y$	Y	$\bar{Y}$	
X	0	2000	2000
$\bar{X}$	6000	2000	8000
	6000	4000	10000

(4) Incompatibilité : $\chi^2 = 3750$, $M_{GK} = -1$

$X \setminus Y$	Y	$\bar{Y}$	
X	3000	0	3000
$\bar{X}$	3000	4000	7000
	6000	4000	10000

(5) Implication logique
 $\chi^2 = 2857$, $M_{GK} = 1$

Figure 1. Cinq situations de référence

Références

- [1] S. GUILLAUME, << Traitement des données volumineuses. Mesures et algorithmes d'extraction des règles d'association et règles ordinales >>, *Universté de Nantes, Phd*, 2000.
- [2] X. WU, C. ZHANG, S. ZHANG, << Efficient mining of both positive and negative association rules >>, *ACM Transactions on information Systems*, 381–405, 2004.
- [3] J. DIATTA, H. RALAMBONDRAINY, A. TOTOHASINA, << Towards a Unifying Probabilistic ImplicativeNormalized Quality Measure for Association Rules >>, *Quality Measures in Data Mining book*, 10, 237–250, 2007.
- [4] S. FERRÉ, << Systèmes d'information logique : un paradigme logico-contextuel pour interroger, naviguer et apprendre >>, *Université de Rennes I, Phd*, 2006.
- [5] R. AGRAWAL, T. IMIELINSKI, A. SWAMI, << Mining association rules between sets of items in large databases >>, *Proc. of the ACM SIGMOD International Conference on Management of Data*, editor P. Buneman and S. Jajodia, Washington, U.S.A., 1993.
- [6] I.C. LERMAN, Classification et analyse ordinaire des donnés >>, *Dunod*, 1984.
- [7] J. BLANCHARD, F. GUILLET, H. BRIAND, R. GRAS, << Assessing rule with a probabilistic measure of deviation from equilibrium >>, *Proc. of 11th International Symposium on Applied Stochastic Models and Data Analysis ASMDA, ENST, Brest, France*, n° 74, 191–200, 2005.