

H. BADRAN

**Trace minima et information mutuelle pour un
tableau de correspondance où manque une donnée**

Les cahiers de l'analyse des données, tome 17, n° 2 (1992),
p. 189-196

http://www.numdam.org/item?id=CAD_1992__17_2_189_0

© Les cahiers de l'analyse des données, Dunod, 1992, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

TRACE MINIMA ET INFORMATION MUTUELLE POUR UN TABLEAU DE CORRESPONDANCE OÙ MANQUE UNE DONNÉE

[INF. MANQ.]

H. BADRAN*

1 Introduction: estimation des données manquantes suivant un critère global et choix d'un critère

Un récent travail, (cf. K. Ben SALEM, [TRAC. MANQ.], in *CAD*, Vol. XVII, n°1, 1992), propose d'estimer les données qui manquent dans un tableau de contingence d'après un critère global qui est de rendre minima la trace de la correspondance définie par le tableau complété.

Or il est bien connu (cf. [Inf. Tab.], in *L'Analyse des Données*, TIB, n°5) que, pour une loi f_{IJ} , la trace, (ou distance du χ^2 entre la loi f_{IJ} et le produit, $f_I \times f_J$, de ses marges), est, à un coefficient près, osculatrice à l'information mutuelle entre I et J; l'opinion commune des spécialistes étant, nous semble-t-il, que l'information mutuelle est la notion de référence; tandis que la trace n'est qu'une approximation qu'on adopte parce qu'elle conduit à des calculs simples et s'interprète comme l'inertie d'un nuage de points situé dans un espace euclidien.

C'est pourquoi nous avons tenté de substituer au critère de la *trace* minima un critère de l'*information mutuelle* minima. Il est apparu que ce nouveau critère ne pouvait convenir, même dans le cas le plus simple où il manque une seule donnée. Ce qui conduit à revoir la comparaison entre trace et information mutuelle; la conclusion étant qu'en dehors du domaine d'osculature, au voisinage de l'indépendance entre I et J, c'est la trace qui exprime le plus fidèlement le lien entre I et J.

Au §2, on calcule l'information mutuelle $mut(x)$, pour une loi t_{IJ} de la forme :

$$t_{IJ} = (1 - x) f_{IJ} + x s_{IJ} ;$$

(*) Université Libanaise.

où f_{II} est une loi pour laquelle (i^0, j^0) a la masse nulle; et s_{II} désigne la loi de masse 1 concentrée au point (i^0, j^0) .

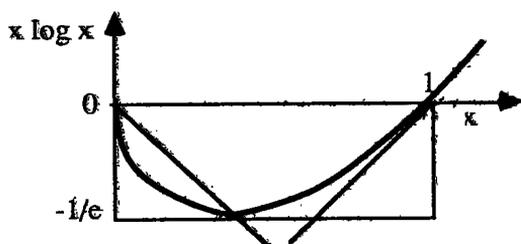
Au §3, on compare la fonction $\text{mut}(x)$ trouvée au §2 avec $\text{trac}(x)$, calculée dans [TRAC. MANQ] pour la même loi t_{II} , dépendant de x .

Au §4 on complète l'étude de $\text{mut}(x)$; la conclusion étant que l'information mutuelle ne convient aucunement comme critère pour reconstituer des données manquantes.

2 Calcul de l'information mutuelle pour une loi binaire combinaison d'une masse ponctuelle et d'une loi donnée

2.1 Rappel de la définition de l'information mutuelle

Sans entreprendre un exposé de théorie de l'information tel que celui donné dans [Inf. Tab.], nous fixerons les notations du présent article en rappelant des formules.



Si $p_I = \{p_i \mid i \in I\}$ est une loi de probabilité sur I , on définit l'information, ou néguentropie, comme une quantité positive ou nulle associée à p_I par la formule:

$$H_2(p_I) = \sum \{-p_i \log_2(p_i) \mid i \in I\};$$

en particulier, si $\text{card} I = 2^n$, et que les p_i sont tous égaux, $H_2(p_I) = n$. L'éventualité $H_2(p_I)$ ne se réalise que si tous les p_i sont nuls, à l'exception d'un seul qui vaut 1.

Dans tous les cas, le calcul de l'information conduit à un résultat fini et bien déterminé, même si certains des f_i sont nuls, du fait que, comme le rappelle le graphique, $(x \log x)$ tend vers zéro quand x tend vers zéro par valeur positive.

Dans la suite, pour simplifier les calculs, nous utiliserons, sous le nom d'information la quantité $H(p_I)$ définie ci-après:

$$H(p_I) = \log(2) H_2(p_I) = \sum \{-p_i \log(p_i) \mid i \in I\}$$

À une loi binaire p_{IJ} , dont les lois marginales sont p_I, p_J , on associe une information mutuelle, quantité positive ou nulle définie par la formule:

$$H_2(p_{IJ} ; p_I \times p_J) = H_2(p_I) + H_2(p_J) - H_2(p_{IJ}) ;$$

ici encore, on introduit une formule simplifiée:

$$Hm(p_{IJ}) = \log(2) H_2(p_{IJ} ; p_I \times p_J) = H(p_I) + H(p_J) - H(p_{IJ}) ;$$

Hm ne s'annule que dans le cas d'une loi produit; de plus, si l'on procède à des cumuls de blocs de lignes ou de blocs de colonnes, Hm ne peut que décroître; et il décroît strictement, excepté dans le cas où le cumul ne porte que sur des blocs de lignes ou de colonnes proportionnelles entre elles. Ceci correspond au principe d'équivalence distributionnelle qui vaut pour les calculs d'inertie.

Comme l'information, l'information mutuelle est bien définie même si certaines masses sont nulles. Au contraire, d'après [TRAC. MANQ.] §2.1, la trace d'un tableau comportant des zéros ne se calcule que par un passage à la limite dont le résultat dépend des valeurs limite de certains rapports de masse.

2.2 Calcul de l'information mutuelle dans un cas particulier

Revenons au cas particulier annoncé au §1:

$$t_{IJ} = (1-x) f_{IJ} + x s_{IJ} \quad ; \quad \text{mut}(x) = Hm(t_{IJ}) \quad ;$$

où f_{IJ} est une loi pour laquelle (i°, j°) a la masse nulle; et s_{IJ} désigne la loi de masse 1 concentrée au point (i°, j°) .

Il est commode de noter:

$$f_{i^\circ} = a \quad ; \quad f_{j^\circ} = b \quad ; \quad c = 1 - (a + b) \quad .$$

On calcule H pour les lois marginales de t_{IJ} et pour t_{IJ} . Il vient, en comptant d'abord à part le terme en i° :

$$\begin{aligned} -H(t_{IJ}) &= (x+(1-x)a) \log(x+(1-x)a) + \sum\{(1-x)f_i \log((1-x)f_i) \mid i \in I, i \neq i^\circ\} \\ &= (x+(1-x)a) \log(x+(1-x)a) - ((1-x)a) \log((1-x)a) \\ &\quad + \sum\{(1-x)f_i \log((1-x)f_i) \mid i \in I\} \\ &= (x+(1-x)a) \log(x+(1-x)a) - ((1-x)a) \log((1-x)a) \\ &\quad + (1-x) \log(1-x) - (1-x) H(f_I) \quad ; \end{aligned}$$

on a de même:

$$\begin{aligned} -H(t_{IJ}) &= (x+(1-x)b) \log(x+(1-x)b) - ((1-x)b) \log((1-x)b) \\ &\quad + (1-x) \log(1-x) - (1-x) H(f_J) \quad ; \end{aligned}$$

on notera que, pour $x=1$, $H(t_I)$ et $H(t_J)$ sont bien définis et nuls.

Pour calculer $H(t_{IJ})$ on procède de même; on a, sans faire cas du terme en (i°, j°) qui dans la somme Σ est nul:

$$\begin{aligned} -H(t_{IJ}) &= x \log(x) + \sum \{ (1-x) f_{ij} \log((1-x) f_{ij}) \mid i \in I; j \in J \} \\ &= x \log(x) + (1-x) \log(1-x) + (1-x) \sum \{ f_{ij} \log(f_{ij}) \mid i \in I; j \in J \} \\ &= x \log(x) + (1-x) \log(1-x) - (1-x) H(f_{IJ}) \quad ; \end{aligned}$$

où on a tenu compte de ce que $\sum \{ f_{ij} \mid i \in I; j \in J \} = 1$.

On peut maintenant achever le calcul de $\text{mut}(x) = \text{Hm}(t_{IJ})$:

$$\begin{aligned} \text{Hm}(t_{IJ}) &= H(t_I) + H(t_J) - H(t_{IJ}) \\ &= (((1-x)a) \log((1-x)a) - (x+(1-x)a) \log(x+(1-x)a)) \\ &\quad + (((1-x)b) \log((1-x)b) - (x+(1-x)b) \log(x+(1-x)b)) \\ &\quad + x \log(x) - (1-x) \log(1-x) \\ &\quad + (1-x) \text{Hm}(f_{IJ}) \quad ; \end{aligned}$$

pour $x=0$, $t_{IJ} = f_{IJ}$ et $\text{Hm}(t_{IJ}) = \text{Hm}(f_{IJ})$; pour $x=1$, la loi t_{IJ} est concentrée sur la paire (i°, j°) et $\text{Hm}(t_{IJ})$ est nul, comme $H(t_{IJ})$, $H(t_I)$ et $H(t_J)$.

Il apparaît que le minimum de $\text{mut}(x)$ est toujours zéro, valeur prise au moins une fois pour $x=1$; ce qui semble déjà interdire le choix de mut comme critère de reconstitution d'une donnée manquante. Mais avant de préciser, au §4, la variation de $\text{mut}(x)$, nous comparerons, au §3, la forme paramétrique de $\text{mut}(x)$ à celle de $\text{trac}(x)$.

3 Formes paramétriques de la trace et de l'information mutuelle pour une loi binaire combinaison d'une masse ponctuelle et d'une loi donnée

Conservons les notations du §2. La fonction $\text{mut}(x)$ ne dépend de la loi f_{IJ} que par l'intermédiaire de 3 paramètres seulement: $a = f_{i^\circ}$; $b = f_{j^\circ}$; $\text{Hm}(f_{IJ})$.

$$f1 = \begin{array}{|c|c|c|} \hline 0 & a & \\ \hline & & \\ \hline b & c & \\ \hline \end{array} ; \quad f2 = \begin{array}{|c|c|c|} \hline 0 & a & 0 \\ \hline b & 0 & 0 \\ \hline 0 & 0 & C \\ \hline \end{array}$$

On a dit au §2.1 que, si l'on procède à des cumuls de blocs de lignes ou de blocs de colonnes, Hm ne peut que décroître: le minimum de $\text{Hm}(f_{IJ})$, pour $\{a, b\}$ fixés, est donc atteint avec le tableau 2×2 , $f1_{IJ}$; tandis que le maximum de $\text{Hm}(f_{IJ})$ peut être arbitrairement grand, à condition que c ne soit pas nul, comme on le voit sur le schéma $f2_{IJ}$, où la lettre C désigne un bloc carré $n \times n$, dont les termes extradiagonaux sont nuls, les termes diagonaux valant c/n .

Pour la trace $\text{crit}(x)$ de t_{IJ} , on a, d'après [TRAC. MANQ.] §4, la formule:

$$\text{trac}(x) = x^2 / ((a+x(1-a)) (b+x(1-b))) + (1-x)h / (a+x(1-a)) + (1-x)g / (b+x(1-b)) + Y - 1 ;$$

où $Y = \sum\{(f_{ij})^2 / (f_i f_j) \mid i \neq i^0, j \neq j^0\}$ et h et g désignent des sommes, satisfaisant aux inégalités rappelées ci-après:

$$h = \sum\{(f_{i^0j})^2 / f_j \mid j \neq j^0\} \leq a \quad ; \quad g = \sum\{(f_{ij^0})^2 / f_i \mid i \neq i^0\} \leq b \quad .$$

La formule $\text{trac}(x)$ comprend 5 paramètres $\{a, b, h, g, Y\}$: procédant comme pour $\text{mut}(x)$, nous nous proposons de considérer les valeurs possibles pour $\{h, g, Y\}$, $\{a, b\}$ étant fixés.

L'égalité $h=a$ n'est réalisée que si les f_{ij} non nuls (autres que ceux situés dans la ligne i^0 ou la colonne j^0) sont tels que $f_{i^0j}=0$; en ce cas, en effet, pour chacun des termes non nuls de la somme \sum qui donne h , $f_{i^0j}=f_j$. Afin de préciser l'intervalle des valeurs possibles pour h , convenons de noter:

$$\forall j \in J : c_j = \sum\{f_{ij} \mid i \neq i^0\} \quad ; \quad a_j = f_{i^0j} \quad ;$$

avec ces notations, la somme \sum qui donne h s'écrit:

$$h = \sum\{(a_j)^2 / (a_j + c_j) \mid j \neq j^0\} \quad ;$$

pour trouver le minimum de h , on tient compte des contraintes $a = \sum\{a_j \mid j \neq j^0\}$ et $c = \sum\{c_j \mid j \neq j^0\}$; on trouve, par différenciation, les conditions de stationarité:

$$\forall j \in J - \{j^0\} : ((a_j)^2 / (a_j + c_j)^2) = cta \quad ; \quad ((c_j)^2 / (a_j + c_j)^2) = ctc \quad ;$$

où cta et ctc désignent deux constantes indépendantes quelconques; on doit donc avoir, compte tenu des contraintes:

$$\forall j \in J - \{j^0\} : (a_j) / (a_j + c_j) = a / (a + c) \quad ;$$

d'où le minimum de h ; et, finalement les inégalités:

$$a^2 / (a + c) \leq h \leq a \quad ; \quad a / (a + c) \leq h / a \leq 1 \quad ;$$

où $(a + c)$ peut être remplacé par sa valeur $(1 - b)$.

Pour les valeurs de g on a les inégalités analogues:

$$b^2 / (b + c) \leq g \leq b \quad ; \quad b / (b + c) \leq g / b \leq 1 \quad ;$$

où $(b + c)$ peut être remplacé par sa valeur $(1 - a)$.

$$f1 = \begin{vmatrix} 0 & a \\ b & c \end{vmatrix} ; \quad f3 = \begin{vmatrix} 0 & a & 0 \\ b & c' & b' \\ 0 & a' & c'' \end{vmatrix}$$

Il est clair que, pour le tableau carré 2×2 , $f1_{II}$, déjà considéré à propos de $\text{mut}(x)$, les deux rapports (h/a) et (g/b) prennent, simultanément, leur valeur minima. Plus généralement, pour des valeurs convenables de $\{a', b', c', c''\}$, quantités positives dont la somme est c , le tableau carré 3×3 , $f3_{II}$, donnera aux deux rapports tout couple de valeurs compatibles avec les inégalités démontrées.

De façon précise, soit $\{A, B\}$ deux nombres compris entre 0 et c ; imposons les conditions:

$$h / a = a / (a + A) \quad ; \quad g / b = b / (b + B) \quad ;$$

il vient pour $\{a', b', c'\}$:

$$0 \leq b' + c' = B \leq c \quad ; \quad 0 \leq a' + c' = A \leq c \quad ;$$

on choisit pour la somme $(a' + b' + c')$ une valeur quelconque satisfaisant à la condition:

$$\sup(A, B) \leq (a' + b' + c') \leq \inf(A + B, c) \quad ,$$

et les valeurs de $\{a', b', c', c''\}$ en résultent immédiatement.

Reste Y : comme cette quantité n'intervient que comme une constante ajoutée à l'expression de $\text{trac}(x)$, nous n'en étudions pas en détail le domaine de variation; nous bornant à montrer que Y peut prendre une valeur arbitrairement grande si on fixe $\{a, b, A, B\}$, avec $0 < A < c$ et $0 < B < c$.

$$f4 = \begin{vmatrix} 0 & a & 0 & 0 \\ b & c' & b' & 0 \\ 0 & a' & 0 & 0 \\ 0 & 0 & 0 & c \end{vmatrix}$$

En effet, en bref, pour construire le tableau $f4_{II}$ proposé ici, on procède comme pour déterminer $f3_{II}$, en posant $(a' + b' + c') = \sup(A, B)$ pour fixer $\{a', b', c'\}$; (e.g., si $B \leq A$, on pose $b'=0, c'=B, a'=A-B$); et la masse restante, $c''=c-\sup(A, B)$, est attribuée à un bloc carré diagonal $n \times n$, C , dont la contribution à Y est n , qu'on peut prendre arbitrairement grand.

Ici apparaissent entre $\text{mut}(x)$ et $\text{trac}(x)$ deux différences essentielles: d'une part, $\text{trac}(x)$ dépend de 5 paramètres, au lieu de 3 pour $\text{mut}(x)$; mais, d'autre part, dans $\text{trac}(x)$, la quantité Y , susceptible d'être arbitrairement grande, n'intervient qu'additivement; tandis que, dans $\text{mut}(x)$, on a le terme non

constant $(1-x) H_m(f_{IJ})$, dont le coefficient $H_m(f_{IJ})$ peut être arbitrairement grand. C'est ce terme qui interdit que $\text{mut}(x)$ présente, dans tous les cas, un minimum, fût-il local, autre que la valeur 0, prise pour $x=1$.

4 Variation de l'information mutuelle pour une loi binaire combinaison d'une masse ponctuelle et d'une loi donnée

Reprenons la formule:

$$\begin{aligned} \text{mut}(x) = & \left(((1-x)a) \log((1-x)a) - (a+(1-a)x) \log(a+(1-a)x) \right) \\ & + \left(((1-x)b) \log((1-x)b) - (b+(1-b)x) \log(b+(1-b)x) \right) \\ & + x \log(x) - (1-x) \log(1-x) \\ & + (1-x) H_m(f_{IJ}) \quad ; \end{aligned}$$

compte tenu de ce que la fonction $x \log(x)$ a pour dérivée $(1 + \log(x))$, il vient:

$$\begin{aligned} \text{mut}'(x) = & \log(x) + \log(1-x) - H_m(f_{IJ}) \\ & - \left(a \log((1-x)a) + (1-a) \log(a + (1-a)x) \right) \\ & - \left(b \log((1-x)b) + (1-b) \log(b + (1-b)x) \right) \quad ; \end{aligned}$$

et pour la dérivée seconde:

$$\begin{aligned} \text{mut}''(x) = & (1/x) - (c/(1-x)) \\ & - \left(((1-a)^2/(a + (1-a)x)) + ((1-b)^2/(b + (1-b)x)) \right) \\ = & \text{Num}(x) / \text{Den}(x) \quad ; \end{aligned}$$

où le numérateur et le dénominateur sont les deux polynômes:

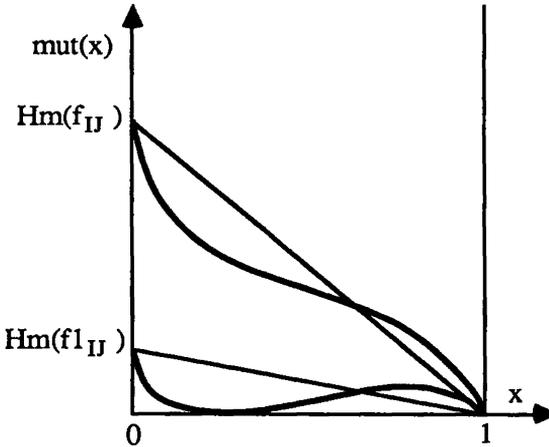
$$\begin{aligned} \text{Num}(x) = & (1-a)(1-b)x^2 - ab \quad ; \\ \text{Den}(x) = & (a + (1-a)x)(b + (1-b)x)(x^2 - x) \end{aligned}$$

Sur l'intervalle $(0, 1)$, $\text{Num}(x)$ s'annule pour $x = x^\mu = (ab/((1-a)(1-b)))^{1/2}$; et $\text{Den}(x)$ est constamment négatif; au voisinage des extrémités de l'intervalle $(0, 1)$, on a:

$$x \approx 0 : \text{mut}''(x) \approx 1/x \quad ; \quad 1-x \approx 0 : \text{mut}''(x) \approx -c/y \quad ;$$

La dérivée première, $\text{mut}'(x)$ vaut $(-\infty)$ pour $x=0$; elle atteint son maximum pour $x=x^\mu$; décroît ensuite vers $(-\infty)$ pour $x=1$. Les paramètres $\{a, b\}$ étant fixés, le signe du maximum dépend de la valeur de $H_m(f_{IJ})$, valeur qui peut être arbitrairement grande, et figure avec le signe '-' dans l'expression de $\text{mut}'(x)$. La plus faible valeur possible de $H_m(f_{IJ})$, et donc la plus forte valeur du maximum, $\text{mut}'(x^\mu)$, de $\text{mut}'(x)$, s'obtiennent dans le cas du tableau 2×2 , f_{1IJ} , considéré au début du §3. Cette plus forte valeur possible de $\text{mut}'(x^\mu)$ est

positive; mais, pour une valeur suffisamment élevée de $Hm(f_{IJ})$, $mut'(x^\mu)$ sera négatif, et, par conséquent, $mut'(x)$ sera négatif sur tout l'intervalle $(0, 1)$.



La figure ci-dessus correspond au cas où $a = b = c = 1/3$; on a considéré la loi f_{1IJ} ainsi qu'un autre cas pour lequel il n'y a pas d'extremum pour $mut(x)$ entre $x=0$ et $x=1$.

Si $mut'(x^\mu)$ est positif, $mut(x)$ décroît à partir de la valeur $mut(0)=Hm(f_{IJ})$ jusqu'à un minimum (qui est nul dans le cas du tableau f_{1IJ} , et correspond à $x^m=(x^\mu)^2$); puis, (après avoir passé par une inflexion pour $x=x^\mu$), croît jusqu'à un maximum local; et enfin décroît jusqu'à zéro quand x tend vers 1. Si $mut'(x^\mu)$ est négatif, $mut(x)$ décroît constamment, à partir de la valeur $mut(0)=Hm(f_{IJ})$ jusqu'à $mut(1)=0$.

C'est pourquoi nous avons annoncé, à la fin du §3, que le terme $(1-x) Hm(f_{IJ})$ interdit que $mut(x)$ présente, dans tous les cas, un minimum, fût-il local, autre que la valeur 0, prise pour $x=1$.

Références bibliographiques

[Inf. Tab], in *L'Analyse des Données*, TIB n°5.

[TRAC. MANQ.], in *CAD*, Vol. XVII, n°1, (1992).