S. CHAHDOURA

Étude sur un cas modèle de questionnaire du double recadrage des notes suivant l'équation personnelle

Les cahiers de l'analyse des données, tome 16, nº 3 (1991), p. 305-312

<http://www.numdam.org/item?id=CAD_1991__16_3_305_0>

Les Cahiers de l'Analyse des Données
Vol XVI - 1991 - n° 3 - pp. 305-312.

ÉTUDE SUR UN CAS MODÈLE DE QUESTIONNAIRE DU DOUBLE RECADRAGE DES NOTES SUIVANT L'ÉQUATION PERSONNELLE

[DOUB. REC. PERS.]

*S. CHAHDOURA**

1 Rappel: le recadrage des notes

On sait que, dans de nombreuses enquêtes, figure un bloc de questions dont chacune demande aux sujets de marquer sur une échelle le degré d'approbation qu'ils accordent à une phrase, ou leur position entre deux phrases exprimant des attitudes opposées. Parfois, l'échelle est présentée comme une ligne horizontale sans repère distinctif; parfois l'échelle est divisée par des traits transversaux, voire graduée numériquement; dans certains cas, une suite de qualificatifs (e.g. de “tout à fait opposé” à “tout à fait d'accord”) réalise verbalement les degrés de l'échelle.

Cette présentation a l'avantage d'offrir au sujet le moyen de s'exprimer avec quelques nuances, au lieu de lui imposer un choix, souvent embarassant, entre le ‘Oui’ et le ‘Non’. Mais de la liberté qui lui est laissée, le sujet n'use pas seulement pour répondre à la question même, mais aussi pour manifester telle tendance personnelle à fuir les extrémités, ou à se décaler complaisamment vers l'approbation.

On s'affranchit de ce biais, d'une part, en concevant des questionnaires où alternent aléatoirement des questions de polarités diverses qui imposent à un sujet conséquent de marquer des réponses à droite et à gauche de l'échelle; et, d'autre part, en corrigeant les notes données par chaque sujet, pour faire disparaître les tendances que manifestent la ditribution statistique de celles-ci. En bref, par une formule semi-linéaire, les notes sont recadrées entre -1 et +1, le zéro correspondant à la moyenne des notes initiales: c'est ce qu'on appelle: corriger suivant l'*équation personnelle*.

Les réponses, ainsi recadrées, à chaque question q peuvent être codées barycentriquement suivant trois modalités {q-, q=, q+}, avec pour valeurs pivot

(*) Étudiant en Doctorat à l'Université Pierre et Marie Curie.

{-1, 0, +1}; mais parce que, pour certaines questions qui suscitent une approbation ou un rejet quasi général, l'une des modalités extrêmes, q- ou q+, peut être très légère (voire vide), on a, de plus, proposé de choisir des valeurs pivot propres à chaque question: {min, moy, Max}; valeurs qui sont, respectivement le minimum, la moyenne et le maximum des notes (recadrées) inscrites dans la colonne q. D'où un codage suivant trois modalités {q<, q≈, q>}, qu'on peut considérer comme résultant d'un double recadrage.

L'objet du présent travail est d'apprécier, dans un cas modèle, en termes mathématiques, l'effet du deuxième recadrage.

2 Un modèle continu unidimensionnel

On sait que les modèles continus sont souvent plus faciles à traiter que les modèles finis, discrets: il en est notamment ainsi dans le cas classique de l'échelle de GUTTMAN.

Nous supposons ici que l'ensemble I des individus est assimilé à une distribution de masse, de densité constante (1/2), sur le segment (-1, +1). Les questions sont réparties en deux ensembles, A et B, l'un et l'autre identifié à ce même segment. La réponse d'un sujet *i* à une question, *a* ou *b*, est notée comme un nombre, compris entre -1 et +1, et donné par les formules:

$$ka(i, a) = (-1) + ((1 + a)(1 + i) / 2) \quad ;$$

$$kb(i, b) = (+1) - ((1 + b)(1 + i) / 2) \quad .$$

Il faut justifier ces formules: d'une part, en proposant une interprétation vraisemblable; d'autre part, en montrant la simplicité mathématique du modèle.

Pour une question *a* donnée, la réponse k(i, a) est distribuée uniformément entre le minimum absolu (-1), qui est la réponse du sujet (i = -1), et *a* qui est le maximum propre à la question et la réponse du sujet (i = +1). On peut dire que toutes les questions de A ont, au fond, la même signification; mais que, du fait de l'expression, les sujets sont d'autant plus réticents à fournir une réponse positive que *a* est plus proche du minimum (-1). Ainsi chaque question a son intervalle propre (-1, a), où les sujets se placent comme ils le sont sur I, mais avec un changement d'échelle.

Pour les questions de B, on a la disposition symétrique de celle qui vaut pour A. Pour *b* donné, la réponse k(i, b) est distribuée uniformément entre le maximum absolu (+1), réponse du sujet (i = -1), et -b, qui est le minimum propre à la question, réponse de (i = +1). On peut dire que les questions B expriment toutes le contraire de ce que propose A; mais tandis qu'avec A la forme est telle que les sujets évitent de se prononcer jusqu'à l'extrémité (+1), avec B, au contraire, c'est le rejet absolu, l'extrémité (-1) qu'on évite.

A des questions *a* et *b* de même abscisse *x*, un sujet *i* fournit des réponses opposées: ka(i, x) = -kb(i,x). Les réponses au questions de A données par un même sujet *i* sont distribuées uniformément sur l'intervalle (-1, i); de même, les réponses aux questions de B sont distribuées uniforméméent sur (-i, +1). Si (i = 1), les deux distributions s'étendent sur tout le segment (-1, +1): le sujet utilise au maximum les possibilités d'expression qui lui sont offertes. Si (i = -1), au contraire, il n'y a que la réponse (-1) à A et la réponse (+1) à B.

Selon que l'abscisse *i* est positive ou négative, les distributions des réponses à A et B empiètent ou non. Mais, quel que soit *i*, l'ensemble des réponses à A et B a pour minimum (-1) pour maximum (+1), pour moyenne 0: il n'y a donc pas lieu d'effectuer un premier recadrage suivant l'équation personnelle. Le modèle se prête donc à l'étude de l'effet pur et simple du recadrage des colonnes.

3 Analyse du modèle après recadrage des colonnes

Après recadrage des colonnes, toutes les questions, celles de A comme celles de B deviennent identiques; le codage barycentrique du modèle continu aboutit donc à un simple tableau à 3 colonnes dont il sera facile d'achever l'analyse. Au contraire, sans ce recadrage, le nuage des individus est de dimension infinie, ainsi qu'on le montrera au §4. La simplification apportée par le recadrage des colonnes est donc ici manifeste..

On a dit que la réponse à une question *a* est distribuée uniformément entre le minimum (-1) et le maximum (a). Cette uniformité nous permet de raisonner en transformant linéairement les valeurs au lieu de déplacer les pivots. Après recadrage de la colonne entre (-1) et (+1), on a des valeurs distribuées uniformément sur (-1, +1) et la réponse du sujet (i) sur les trois modalités {a<, a≈, a>}, définies par les pivots {-1, 0, +1}, est donnée par les formules:

si (i >0) alors {0, 1-i, i} ; si (i<0) alors {-i, 1+i, 0} .

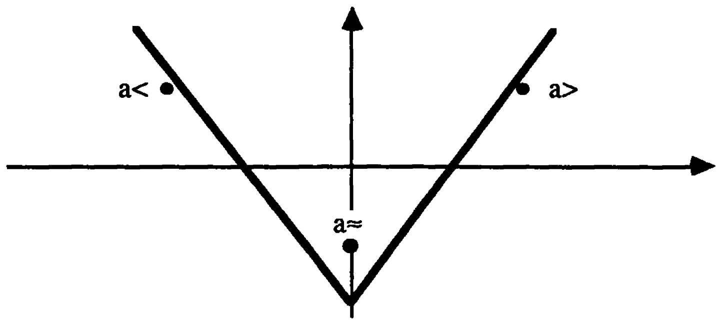
Pour une question *b*, le résultat ultime est le même: l'intervalle (-b, +1) des valeurs de la colonne est recadré sur (-1, +1); et l'on a sur {b<, b≈, b>} les mêmes valeurs, en fonction de i, que sur {a<, a≈, a>}.

Avec 3 colonnes, l'espace support des nuages est un plan. Il résulte de la linéarité des formules que le nuage N(I) se compose de deux segments de droite, correspondant aux sous-intervalles (-1, 0) et (0, +1) de I= (-1, +1). Les extrémités des segments sont les trois profils:

i- = {1, 0, 0} ; i0 = {0, 1, 0} ; i+ = {0, 0, 1} ;

Parce que les colonnes {<, ≈, >} ont pour poids respectifs {1/4, 1/2, 1/4}, on a les distances distributionnelles:

d^2(i-, i+) = 8 ; d^2(i-, i0) = d^2(i+, i0) = 6 .



De ces distances, en faisant choix d'une orientation des axes, on déduit les valeurs des facteurs:

$F_1(i\text{-}) = -\sqrt{2}$; $F_1(i\text{-}) = 0$; $F_1(i+) = +\sqrt{2}$;

$F_2(i\text{-}) = F_2(i+) = +1$; $F_2(i0) = -1$;

pour les autres valeurs de i, les formules se complètent par linéarité sur les deux segments:

$F_1(i) = i\sqrt{2}$; $F_2(2) = (2\,|\,i\,|) - 1$.

On calcule simplement les facteurs sur J = {a<, a≈, a>} en passant par les valeurs propres λ_1 et λ_2 comme l'inertie d'un segment de masse 1 et de longueur respective $2\sqrt{2}$ et 2, relativement à son centre de gravité:

$\lambda_1 = 2/3$; $\lambda_2 = 1/3$;

d'où, du fait de la symétrie propre à J et de son inertie sur les axes:

$F_1(a<) = -2/(\sqrt{3})$; $F_1(a\approx) = 0$; $F_1(a>) = +2/(\sqrt{3})$;

$F_2(a<) = F_2(a>) = 1/(\sqrt{3})$; $F_2(a\approx) = -1/(\sqrt{3})$.

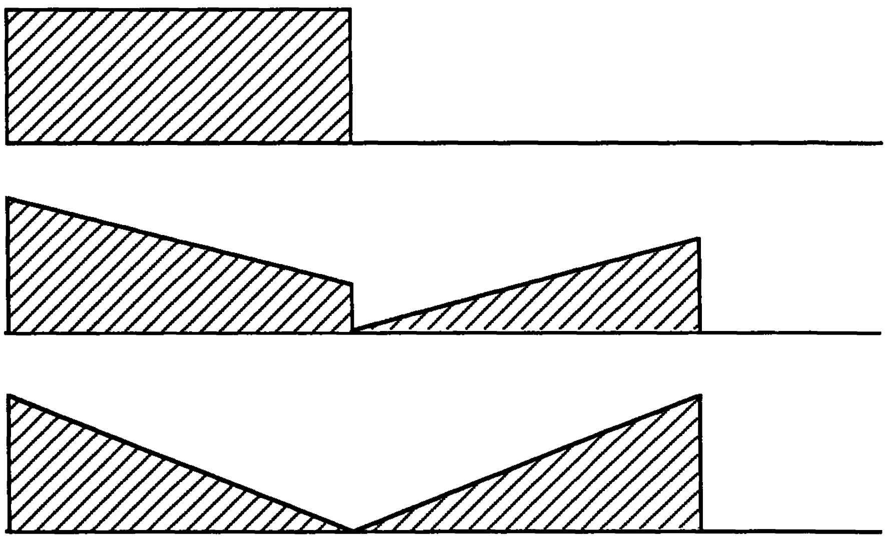
4 Les nuages N(I) et N(J) sans recadrage des colonnes

Puisque l'ensemble Q des questions, ou variables de base, est constitué de deux intervalles, A et B, l'ensemble J des modalités doit comprendre 6 intervalles: {A-, A=, A+, B-, B=, B+}. Mais de la formule:

ka(i, x) = - kb(i,x), il résulte que, pour deux questions ax et bx de même abscisse x, on a, après codage, quel que soit l'individu *i* :

k(i, ax-) = k(i, bx+) ; k(i, ax=) = k(i, bx=) ; k(i, ax+) = k(i, bx-) ;

on se bornera donc à considérer les modalités de réponse aux questions de l'intervalle A.



{A- } ≈ (-1, +1) {A=} ≈ (-1, +1) {A+} ≈ (0, +1)

On a représenté ci-dessus le profil de i, *sur l'ensemble* J, *restreint aux modalités de* A, *pour des valeurs négatives de l'abscisse de* i, *plus précisément, pour* i=-1, *en haut,* i=0, *en bas, et une valeur intermédiaire,* i<0. *On notera que les profils sont nuls sur les modalités* a+.

De plus, tandis que chacun des ensembles A- et A= s'identifie, comme A, à l'intervalle (-1, +1) tout entier, l'ensemble A+ doit être restreint à l'intervalle (0, +1) parce que, pour une question *a* d'abscisse négative, la modalité a> est vide (de masse nulle). En effet, les pivots étant fixés à {-1, 0, +1}, k(i, a-) n'est différent de zéro que si ka(i, a) est positif, ce qui ne peut se produire que si (cf. §2) l'intervalle (-1, a) comprend la valeur 0.

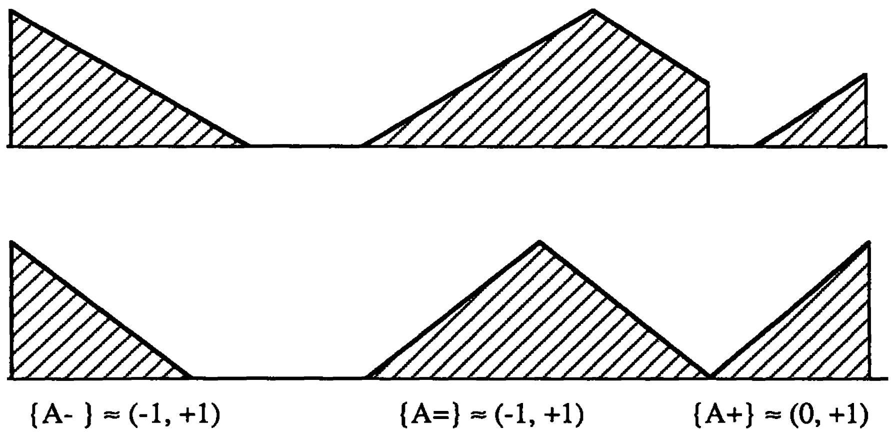
De façon précise, on, pour i<0, les formules simples ci-après, définissant le profil de *i* sur J:

$$\forall\ i \in (-1, 0), \forall\ a \in (-1, +1) : \ ka(i, a) = -1 + ((1+a)\ (1+i) / 2) < 0 \ ;$$

$$k(i, a-) = -\ ka(i, a) \ ; k(i, a=) = 1 + ka(i,a) \ ; k(i, a+) = 0 \ ;$$

c'est d'après ces formules qu'on a dessiné trois profils.

On peut vérifier sur les formules que, comme le suggère la figure, le profil de i<0 est combinaison linéaire des profils correspondant à i=-1 et i=0 avec les coefficients respectifs (-i) et (1+i). À l'intervalle (-1 ≤ i ≤ 0), correspond donc dans l'espace des profils, un segment portant une distribution de masse uniforme.



On a représenté ci-dessus le profil de i, *sur l'ensemble* J, *restreint aux modalités de* A, *pour des valeurs positives de l'abscisse de* i, *plus précisément, pour* i=+1, *en bas et pour une valeur intermédiaire,* i<0, *en haut. On notera que les profils s'étendent aux modalités* a+.

Mais à l'intervalle $(0 < i \leq +1)$ correspond un ensemble de profils tout autrement complexe. Pour chaque (i>0), les formules changent selon que *a* est inférieur ou supérieur à la valeur limite pour laquelle ka(i, a) =0, soit:

$$a(i) = (1 - i) / (1 + i) > 0 \ ;$$

on a les formules:

$$\forall \ i \in (0, +1), \forall \ a \in (a(i), +1) : \ ka(i, a) = -1 + ((1+a)(1+i) / 2) > 0 \ ;$$

$$k(i, a-) = 0 \ ; k(i, a=) = 1 - ka(i,a) \ ; k(i, a+) = ka(i, a) \ ;$$

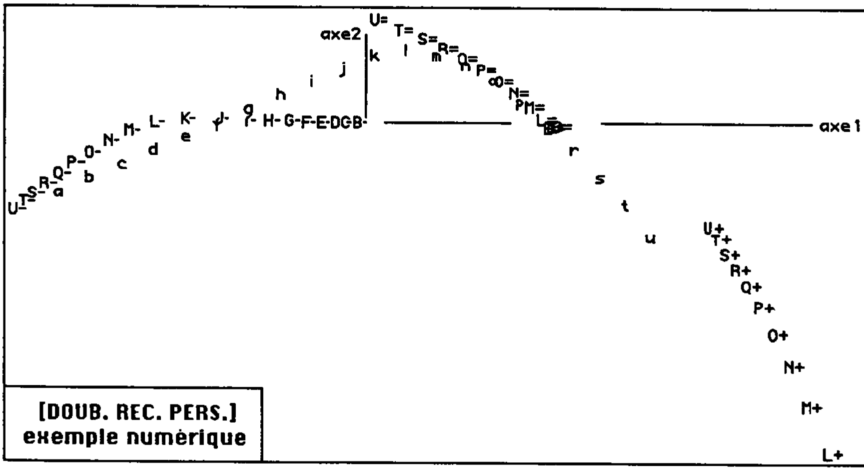
$$\forall \ i \in (0, +1), \forall \ a \in (-1, a(i)) : \ ka(i, a) = -1 + ((1+a)(1+i) / 2) < 0 \ ;$$

$$k(i, a-) = - ka(i, a) \ ; k(i, a=) = 1 + ka(i,a) \ ; k(i, a+) = 0 \ .$$

Le profil de (i>0) est donc formé de trois créneaux. De plus, dans l'espace des profils, il correspond à l'intervalle $(0 < i \leq +1)$ une courbe dont le support, bien loin d'être une droite, est de dimension infinie. Il suffit pour voir cela de considérer la restriction des profils à l'ensemble {A+} des modalités a+. En effet, soit un ensemble fini de *n* individus dont les abscisses positives, rangées dans l'ordre croissant, sont $\{i_1, i_2, \ldots, i_n\}$; on a par conséquent:

$$a(i_n) < a(i_{n-1}) < \ldots \ < a(i_2) < a(i_1) \ ;$$

entre les restrictions des profils il ne peut y avoir de relation linéaire homogène à coefficients non nuls $\{u_1, u_2, \ldots, u_n\}$:



$\forall$ a $\in$ (0, +1) :

$$0 = u_1\, k(i_1, a+) + u_2\, k(i_2, a+) + \ldots + u_{n-1}\, k(i_{n-1}, a+) + u_n\, k(i_n, a+)\ ;$$

car sur l'intervalle ($a(i_n) \leq a < a(i_{n-1})$), le membre de droite de l'égalité se réduit à son dernier terme; lequel est non nul, si u_n est non nul.

On peut, sur un exemple numérique, voir combien la structure est plus complexe que celle obtenue, au §3, après recadrage des colonnes. Il suffit de traiter le cas discret d'un ensemble de 21 individus, {a, b, ..., u} et d'un ensemble de 21 questions {A, B, ..., U}; avec, pour chacun des deux ensembles des abscisses (*i* ou *a*) variant en progression arithmétique de -1 à +1.

L'abscisse zéro correspond à la question K; seules les 10 questions {L, ...,U} admettent 3 modalités non vides (e.g. R-, R=, R+); les questions {B, ...,K} n'ont que deux modalités (e.g. D-, D=). Quant à la question A, sa modalité A= est vide; et A-, ayant un profil plat (même valeur 1 pour tous les individus), se projette à l'origine et a donc été éliminée.

On vérifie, sur le graphique du plan (1,2), l'alignement des individus {a, ..., k} d'abscisse négative ou nulle; les autres individus dessinent une courbe qui prolonge le segment (a, k).

De plus, on constate que les modalités '=' propres aux questions d'abscisse *a* telle que ($-1 < a \leq 0$) sont superposées: en effet, on vérifie par le calcul que les colonnes a= afférentes à ces question ont toutes, sur I, le même profil, qui est proportionnel à (1+i). De même, les modalités '-' de ces mêmes questions sont alignées, ayant toutes sur I un profil qui varie linéairement en fonction de l'abscisse *i*.

5 Conclusion

Le recadrage des colonnes a été introduit afin d'éviter que certaines modalités très légères ne perturbent l'analyse des cas réels du fait de fluctuations d'échantillonnage. On a vu ici un cas modèle où le recadrage des colonnes fait apparaître une structure plane particulièrement simple; tandis que, sans cette transformation, on a une structure complexe de dimension infinie.

Références bibliographiques

A. ALAWIEH: "Cas modèle de l'analyse d'une variable continue unique codée barycentriquement", [MOD. CODE BARY.], in *CAD*, Vol XVI, n° 1, (1991).

J.-P. & F. BENZÉCRI: "Codage linéaire par morceaux et équation personnelle", [ÉQU. PERS.], in *CAD*, Vol XIV, n° 3, (1989).