

J. P. BENZÉCRI

**L'effet de chaînage complet pour des points
de masse un sur la droite en classification
ascendante hiérarchique**

Les cahiers de l'analyse des données, tome 9, n° 2 (1984),
p. 231-238

http://www.numdam.org/item?id=CAD_1984__9_2_231_0

© Les cahiers de l'analyse des données, Dunod, 1984, 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/>

L'EFFET DE CHAINAGE COMPLET POUR DES POINTS DE MASSE UN SUR LA DROITE EN CLASSIFICATION ASCENDANTE HIÉRARCHIQUE

[CHAINAGE DROITE CAH]

par J. P. Benzécri ⁽¹⁾

0 Rappel. Dans un précédent article (cf. [CHAÎNE COMP. CAH] in C.A.D. Vol VII n° 2 pp 189-208) on a donné trois exemples de chaînage complet (i.e. de C.A.H., constituée par agrégation successive de tous les individus à une même classe suivant le schéma (4(3(2(0))))). Ces exemples sont toutefois complexes en ce que, ou bien l'espace ambiant est de dimension élevée (égale au nombre des individus -1), ou bien les individus sont de masse inégale. On propose ici sous forme d'exercices deux exemples où il s'agit d'individus ayant tous la masse 1, et situés sur une droite. Le premier exemple (§§ 1 et 2) est simplement constitué des points d'une progression géométrique de raison $a \geq (1 + \sqrt{2})$. Le deuxième exemple (§§ 3 et 4) vise à disposer les points successifs $x^0, x^1, x^i, x^{i+1}, \dots$, chacun aussi proche que possible du précédent sous la condition qu'on ait un chaînage complet.

Les notations sont les mêmes que dans [CHAÎNE COMP. C.A.H.] ; $v(s)$ désigne le plus proche voisin d'un sommet s ; et le critère d'agrégation est celui de la variance :

$$nv(a, b) = (m_a m_b / (m_a + m_b)) |x^a - x^b|^2$$

où a et b sont deux classes ; m_a, m_b leurs masses et x^a, x^b les abscisses de leur centre de gravité.

1 Premier exemple

1.0 Le nuage donné ; notations : l'ensemble des individus est désigné par la suite $[n]$ des $n+1$ premiers entiers (comptés à partir de 0) :

$$I = [n] = \{0, 1, \dots, n\} ;$$

plus généralement on notera $[p]$ la suite $\{0, 1, \dots, p\}$ pour p entier quelconque. L'espace ambiant au nuage $N([n])$ est la droite réelle R (plus précisément la demi-droite positive R^+), munie de la distance usuelle (valeur absolue de la différence des abscisses) :

$$d(x, y) = |x - y| ; (d(x, y))^2 = (x - y)^2.$$

Les abscisses x^i des points M^i constituent une progression géométrique de raison $a (1 \leq i \leq n)$:

$$x^i = (a)^i.$$

On attribue à tous les points la même masse $m_i = 1$. On considérera éventuellement que n peut être arbitrairement grand.

(1) *Professeur de statistique. Université Pierre et Marie Curie.*

1.1 Calcul des écarts entre points

1.1.1 Calculer $nv(i, i')$ en fonction de i, i' et a .

1.1.2 Déterminer l'indice $v(i)$ du plus proche voisin du point M^i

1.2 Centre de gravité d'une classe: soit p un entier.

1.2.1 Exprimer en fonction de a et p l'abscisse $x^{[p]}$ du centre de gravité de la classe $[p]$.

1.2.2 Calculer : $\eta(p) = x^{[p]} / a^{p+1}$; dire si $\eta(p)$ est inférieur ou supérieur à 1 ; déterminer la limite de $\eta(p)$ pour p tendant vers l'infini.

1.3 Écarts entre points et classes

1.3.1 Calculer $n(v) ([p], p+1) / a^{2p+2} = A(p)$; on mettra $A(p)$ sous la forme $(1 - \eta(p))^q f(p)$ où q est un entier convenablement choisi.

1.3.2 Calculer la limite $A(\infty)$ de $A(p)$ pour p tendant vers l'infini.

1.3.3 Comparer $A(p)$ à $A(\infty)$.

1.4 Comparaison entre écart à un point et écart à une classe

1.4.1 Calculer $nv(p+2, p+1) / a^{2p+2} = B(p)$

1.4.2 A quelle condition a doit-il satisfaire pour que

$$\lim\{A(p)/B(p) \mid p \rightarrow \infty\} \leq 1.$$

Cette condition est-elle vérifiée pour $a = 2, a = 3, a = 4$?

1.4.3 A quelle condition a doit-il satisfaire pour que, quel que soit l'entier p (positif ou nul) le rapport $A(p)/B(p)$ soit strictement inférieur à 1 ?

Existe-t-il un plus petit nombre a satisfaisant à cette condition ?

1.5 Construction de la CAH : on suppose désormais vérifiée la condition trouvée au § 1.4.3.

1.5.1 On suppose, qu'à une étape de la classification on ait pour ensemble de sommets :

$$S = \{[p], [p+1], [p+2], \dots, [n]\}$$

i.e. il y a une classe $[p]$; et les individus de rang supérieur à p ne sont pas agrégés.

Déterminer sous cette hypothèse les plus proches voisins de chaque sommet ?

1.5.2 Représenter l'arbre de la CAH en donnant pour chaque noeud son niveau. Pour la figure et le calcul numérique on pourra se borner au cas $n = 3$ et prendre pour a la plus petite valeur entière compatible avec la condition du § 1.4.3.

Mais on fera une description générale valant pour n quelconque.

2 Explication du premier exemple

2.1.1 $nv(i, i') = (a^i - a^{i'})^2 / 2$

2.1.2 pour $i \geq 1$, $v(i) = i - 1$; $v(0) = 1$; en effet à partir du point M^0 d'abscisse 1, les points s'espacent de plus en plus sur la demi-droite R^+ .

2.2.1 $x^{[p]} = (a^{p+1} - 1)(a - 1)^{-1} (p + 1)^{-1}$. (On applique la formule classique de la sommation d'une série géométrique.

2.2.2 $\eta(p) = (a - 1)^{-1} (1 - a^{-(p+1)})(p + 1)^{-1}$. Il est clair que $\eta(p)$ est inférieur à 1, ce qui revient à dire que $x^p \angle x^{p+1}$, car $x^{[p]}$ est le centre de gravité d'un système de points qui ont tous des abscisses inférieures à celles de x^{p+1} . Quand p tend vers l'infini le premier facteur de $\eta(p)$ reste constant ; le deuxième tend vers 1 ; le troisième tend vers zéro ; donc la limite demandée est zéro.

2.3.1 $A(p) = (1 - \eta(p))^2 ((p+1)/(p+2)) = (1 - \eta(p))^2 (1 - (p+2)^{-1})$

2.3.2 $A(\infty) = 1$

2.3.3 $A(p) \underset{s}{\angle} A(\infty) = 1$

2.4.1 $B(p) = (a-1)^2 / 2$

2.4.2 La limite considérée est $2/(a-1)^{-2}$; cette limite est ≤ 1 , si a est supérieur ou égal à $1 + \sqrt{2}$. Cette condition est vérifiée pour $a = 3$, $a = 4$ mais non pour $a = 2$.

2.4.3 Puisque $A(p) \underset{s}{\angle} A(\infty)$, et que $B(p) = B(\infty) = cte$, il faut et suffit que $(A(\infty)/B(\infty)) \leq 1$, pour que quel que soit p on ait :

$(A(p)/B(p)) \underset{s}{\angle} 1$; et ceci est réalisé si $1 + \sqrt{2} \leq a$.

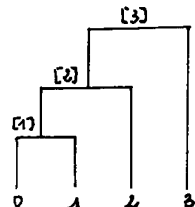
2.5.1 $v([p]) = p + 1$; $v(p+1) = [p]$; $(i-p) \geq 2 \Rightarrow v(i) = i - 1$, pour un individu de rang $\geq p + 2$, il est clair que son plus proche voisin est le point $(i - 1)$ qui le précède à gauche ; car les autres points de masse 1 sont plus éloignés, et le centre $[p]$ également ; or l'écart $nv([p], i)$ est plus fort relativement à la classe $[p]$ de masse $p + 1$, qu'il ne le serait pour un point de masse 1 ayant $x^{[p]}$ pour abscisse (en effet le facteur lié à la masse est $(1/2)$ pour un point isolé et $(p+1)/(p+2)$ pour la classe $[p]$). Quant au plus proche voisin de $p + 1$, c'est $[p]$ et non $p + 2$, parce que l'on a supposé satisfaite la condition trouvée au § 2.4.3.

2.5.2 La C.A.H. présente un effet de chaînage complet : les seules classes sont d'une part les individus isolés, d'autre part les classes proprement dites $[1]$, $[2]$, ..., $[n]$ formées par agrégation progressive des individus à partir de la gauche. Dans le cas $a = 3$ (plus petit entier supérieur à $1 + \sqrt{2}$), on a :

$nv(0, 1) = 2^2 (1/2) = 2$; $x^{[1]} = 2$

$nv([1], 2) = (9-2)^2 (2/3) = 2.7^2 .3^{-1} = 32,66$; $x^{[2]} = 13/3$

$nv([2], 3) = ((13/3) - 27)^2 (3/4) = 2^2 .3^{-1} .17^2 = 385,33$



3 Deuxième exemple

3.0 Notations : On considère sur la demi-droite positive (munie de la distance usuelle) une suite de points dont les abscisses (qui vont en croissant) sont notées :

$$x(0) ; x(1) ; \dots ; x(n) < x(n+1) ; \dots$$

chaque point est affecté de la masse 1 ; et on note $x[n]$ l'abscisse du centre de gravité de l'ensemble $[n] = \{0, 1, \dots, n\}$ des $n+1$ premiers points de la suite. On suppose que $x(0) = 0$; $x(1) = 1$; on cherche à déterminer la suite des abscisses pour satisfaire à certaines conditions relatives aux niveaux d'agrégation.

3.1 Calcul par récurrence de l'abscisse du centre de gravité.

3.1.1 Calculer $x[n+1]$ en fonction de $x[n]$, $x(n+1)$ et n .

3.1.2 Exprimer la différence $x[n+1] - x[n]$ par une formule du type suivant :

$$x[n+1] - x[n] = (x(n+1) - x[n]) / (u + v) ;$$

où u et v sont des nombres constants (indépendants de n) que l'on déterminera.

3.2 Calcul par récurrence de la suite $x(n)$

3.2.1 Soit ϵ un nombre réel positif ou nul. Calculer $x(n+2)$ en fonction de $x[n]$, $x(n+1)$, n et ϵ , de telle sorte que :

$$nv(n+1, n+2) = (1+\epsilon)^2 nv(n+1, [n]).$$

3.2.2 Exprimer la différence $x(n+2) - x(n+1)$ par une formule du type suivant :

$$x(n+2) - x(n+1) = (x(n+1) - x[n]) (1+\epsilon)^P \cdot Q \cdot (1 - (1/(n+S)))^T ;$$

où P , Q , S , T sont des nombres constants (indépendants de n et ϵ) que l'on déterminera.

3.3 Calcul des différences sous forme de produit : Dans la suite on utilisera les notations suivantes :

$$dm(n) = x(n+1) - x[n] ;$$

$$d(n) = x(n+1) - x(n) ;$$

$$d[n] = x[n+1] - x[n] .$$

3.3.1 En soustrayant l'une de l'autre les expressions de $d(n)$ et de $d[n-1]$ trouvées respectivement aux §§ 3.2.2 et 3.1.2, trouver une relation de la forme :

$$dm(n) = dm(n-1) A(n),$$

où $A(n)$ est une fonction de n (et de ϵ) que l'on précisera.

3.3.2 En se souvenant qu'on a posé ($x(0) = 0$, $x(1) = 1$), exprimer $dm(n)$ comme un produit de valeurs de la fonction $A(n)$.

3.4 Expression asymptotique de la fonction $dm[n]$: On suppose désormais pour simplifier les calculs que $\epsilon = 0$.

3.4.1 On demande de donner pour la fonction $A(n)$ un développement valant pour $n \rightarrow \infty$, et de la forme suivante :

$$A(n-1) = K(1 - Ln^{-1} - Mn^{-2} + \dots) ;$$

où K, L, M sont des nombres réels que l'on déterminera.

3.4.2 On met la fonction $dm(n)$ sous la forme d'un produit, comportant une nouvelle fonction $a(n)$ dont on cherche à préciser l'allure pour $n \rightarrow \infty$:

$$dm(n-1) = K^n n^{-L} a(n),$$

où K et L sont des nombres trouvés au § 3.4.1. Utiliser le résultat du § 3.4.1 pour calculer un nombre réel α tel que :

$$a(n) = a(n-1) (1 + \alpha n^{-2} + \dots).$$

3.4.3 Montrer que la fonction $a(n)$ peut se mettre sous la forme :

$$a(n) = a(\infty) (1 - \alpha' n^{-1} + \dots).$$

ou α' est une constante que l'on précisera tandis que le nombre $a(\infty)$ est une valeur limite qu'on ne cherchera pas à préciser.

3.5 Expression asymptotique de l'abscisse $x(n)$: On met la fonction $x(n)$ sous la forme d'un produit, comportant une nouvelle fonction $b(n)$ dont on cherche à préciser l'allure pour $n \rightarrow \infty$:

$$x(n) = K^n n^{-L} b(n),$$

où K et L sont les nombres trouvés au § 3.4.1 et utilisés au § 3.4.2.

3.5.1 Faisant l'hypothèse que $b(n)$ admet un développement de la forme :

$$b(n) = b(\infty) (1 + \beta n^{-1} + \dots)$$

calculer pour $d(n)$ (i.e. $x(n+1) - x(n)$) un développement de la forme :

$$d(n) = K^n n^{-L} b(\infty) C(1 + \gamma n^{-1} + \dots),$$

où C et γ sont des constantes que l'on précisera (en fonction de β).

3.5.2 En utilisant la formule du § 3.2.2, et le développement de $dm(n)$ objet du § 3.4, calculer pour la fonction $d(n)$ un développement de la forme :

$$d(n) = K^n n^{-L} a(\infty) C'(1 + \gamma' n^{-1} + \dots),$$

où C' et γ' sont des constantes numériques que l'on précisera.

3.5.3 En confrontant les résultats des §§ 3.5.2 et 3.5.1, montrer que la forme de $b(n)$ proposée au § 3.5.1 est correcte ; et calculer $b(\infty)$ et β en fonction de $a(\infty)$ et α .

4 Explication du 2-ème exemple

$$4.1.1 \quad x[n+1] = ((n+1)x[n] + x(n+1)) / (n+2) \\ = x[n] + ((x(n+1) - x[n]) / (n+2)) ;$$

$$4.1.2 \quad x[n+1] - x[n] = (x(n+1) - x[n]) / (n+2) ; \text{ d'où } u = 1, v = 2.$$

4.2.1 La condition demandée s'écrit, en revenant à la définition de nv :

$$(x(n+2) - (x(n+1))^{1/2})^2 = (1+\epsilon)^2 (x(n+1) - x[n])^2 ((n+1)/(n+2)) ; \\ x(n+2) - x(n+1) = (x(n+1) - x[n]) (1+\epsilon) 2^{1/2} (1 - (1/(n+2)))^{1/2}$$

$$4.2.2 \text{ D'où } P = 1 ; Q = 2^{1/2} ; S = 2 ; T = 1/2.$$

4.3.1 On combine les relations déjà obtenues :

$$x(n+1) - x(n) = (x(n) - x[n-1]) (1+\epsilon) (2n/(n+1))^{1/2} ; \\ x[n] - x[n-1] = (x(n) - x[n-1]) / (n+1) ;$$

par soustraction, la différence $x(n) - x[n-1]$ apparaît une nouvelle fois (dans le membre de gauche) et il vient :

$$x(n+1) - x[n] = (x(n) - x[n-1]) (1 - (1/(n+1)) + (1+\epsilon)(2n/(n+1))^{1/2})$$

c'est la relation demandée avec :

$$A(n) = 1 - (1/n) + (1+\epsilon)(2n/(n+1))^{1/2}.$$

4.3.2 On a $x[0] = x(0) = 0$; $dm(0) = x(1) - x[0] = 1$; d'où :

$$dm(n) = A(n) dm(n-1) \\ = A(n) \cdot A(n-1) \dots A(1) dm(0) \\ = A(n) \cdot A(n-1) \dots A(1) ;$$

4.4.1 On rappelle la formule :

$$(1-x)^t = 1 - tx + (t(t-1)/2)x^2 + \dots$$

(cette formule vaut pour t constante réelle ou complexe quelconque et $|x| < 1$). On a ici :

$$A(n-1) = 1 - (1/n) + 2^{1/2} ((n-1)/n)^{1/2} \\ = 1 - (1/n) + 2^{1/2} (1 - (1/n))^{1/2} \\ = (1 + \sqrt{2}) - (1 + \sqrt{2}/2) n^{-1} - (\sqrt{2}/8) n^{-2} + \dots \\ = (1 + \sqrt{2}) (1 - (\sqrt{2}/2) n^{-1} - ((2-\sqrt{2})/8) n^{-2} + \dots)$$

$$\text{On a donc : } K = 1 + \sqrt{2} ; L = \sqrt{2}/2 ; M = (2 - \sqrt{2})/8.$$

4.2.2 On utilise la relation.

$$dm(n-1) = A(n) dm(n-2)$$

avec pour la fonction dm la forme proposée :

$$a(n) (1 + \sqrt{2})^n n^{-\sqrt{2}/2} = a(n-1) (1 + \sqrt{2})^{n-1} (n-1)^{-\sqrt{2}/2} \cdot A(n) ;$$

ou, en simplifiant et remplaçant $A(n)$ par son développement :

$$a(n) = a(n-1)((n-1)/n)^{-\sqrt{2}/2} (1 - (\sqrt{2}/2)n^{-1} - ((2-\sqrt{2})/8)n^{-2} + \dots) ;$$

en appliquant à $((n-1)/n)^{-\sqrt{2}/2} = (1 - (1/n))^{-L}$ la formule rappelée pour $(1-x)^t$, on a :

$$\begin{aligned} a(n) &= a(n-1) (1 + Ln^{-1} + (L(L+1)/2)n^{-2} + \dots) (1 - Ln^{-1} - Mn^{-2} + \dots) \\ &= a(n-1) (1 + ((L(L+1)/2) - M - L^2)n^{-2} + \dots) \\ &= a(n-1) (1 + \alpha n^{-2} + \dots) \end{aligned}$$

où : $\alpha = (3\sqrt{2} - 4)/8 > 0$.

4.4.3 Maintenant $a(\infty)$ apparaît comme un produit infini convergent :

$$a(\infty) = a(n) (a(n+1)/a(n)) ; (a(n+2)/a(n+1)) . (a(n+3)/a(n+2)) \dots$$

$$a(\infty) = a(n) (1 + \alpha(n+1)^{-2} + \dots) (1 + \alpha(n+2)^{-2} + \dots) (1 + \alpha(n+3)^{-2} + \dots) \dots$$

le développement du quotient $(a(\infty)/a(n))$ en fonction de n est de la forme $1 + r(n)$, où $r(n)$ est équivalent à :

$$\alpha((n+1)^{-2} + (n+2)^{-2} + \dots)$$

somme dont on a par intégration l'équivalent (α/n) ; d'où :

$$a(\infty) = a(n) (1 + (\alpha/n) + \dots) ;$$

$$a(n) = a(\infty) (1 - (\alpha/n) \dots)$$

on a donc $\alpha' = \alpha = ((3\sqrt{2} - 4)/8) > 0$.

4.5.1 Ayant postulé un développement pour $b(n)$ on peut par différence calculer un développement pour $d(n)$; en identifiant ce développement avec celui qui résulte des calculs du § 4.4, on justifiera la forme postulée pour $b(n)$.

$$\begin{aligned} d(n) &= x(n+1) - x(n) \\ &= b(n+1)K^{n+1}(n+1)^{-L} - b(n)K^n n^{-L} \\ &= K^n n^{-L} (K b(n+1)((n+1)/n)^{-L} - b(n)) \\ &= K^n n^{-L} b(\infty) (K(1 + \beta n^{-1} + \dots)(1 - Ln^{-1} + \dots) - (1 + \beta n^{-1} + \dots)) \\ &= K^n n^{-L} b(\infty) ((K-1) + n^{-1}(K(\beta-L) - \beta) + \dots) \\ &= K^n n^{-L} b(\infty) \sqrt{2} (1 + n^{-1}(\beta - ((1 + \sqrt{2})/2)) + \dots) ; \end{aligned}$$

on a donc : $C = \sqrt{2}$; $\gamma = \beta - ((1 + \sqrt{2})/2)$.

4.5.2 Partant de la formule trouvée au § 4.2, et du développement du § 4.4 on a :

$$\begin{aligned} d(n) &= 2^{1/2} (n/(n+1))^{1/2} dm(n-1) \\ &= K^n n^{-L} a(\infty) \sqrt{2} (1 - \alpha n^{-1} + \dots) (1 - (1/2)n^{-1} + \dots) ; \end{aligned}$$

4.5.3 L'identification avec la formule du § 4.5.1 apparaît possible en posant :

$$b(\infty) = a(\infty) ;$$

$$\beta - ((1 + \sqrt{2})/2) = -\alpha - (1/2) ; \text{ i.e. : } \beta = (4 + \sqrt{2})/8 > 0$$

Le succès de cette identification n'est aucunement fortuit. Ayant démontré au § 4.4 que $d(n)$ est (comme $dm(n)$) équivalent à $K^n n^{-L} \times \text{cte}$, on sait que $x(n)$ qui est la somme jusqu'au rang n de la série de terme général $d(n)$, est aussi de cette forme ; ceci résulte de ce que la fonction $K^x x^{-L}$ et son intégrale indéfinie :

$$\int_1^x K^t t^{-L} dt,$$

sont des infiniment grands du même ordre pour $x \rightarrow \infty$. Il nous suffira de le vérifier pour $K = e$; on a en intégrant par parties

$$\int_1^x e^t t^{-L} dt = [e^t t^{-L}]_1^x + L \int_1^x e^t t^{-L-1} dt ;$$

où l'intégrale du 1-er nombre et celle du 2-ème divergent toutes deux, mais la deuxième qui comporte en plus un facteur $(1/t)$ est d'un ordre de grandeur inférieur à la 1-ère ; ce qui montre que celle-ci est équivalente à $e^x x^{-L}$ pour $x \rightarrow \infty$.

5 Conclusion générale : L'étude du premier exemple (§§ 1 et 2) montre qu'on peut avoir un effet de chaînage complet avec des points dont les abscisses forment une progression géométrique de raison a supérieure ou égale à $1 + \sqrt{2}$. L'étude du deuxième exemple montre qu'il n'est pas possible d'obtenir cet effet avec un ordre de croissance beaucoup moins grand, de façon précise l'optimum est :

$$x(n) \approx (1 + \sqrt{2})^n n^{-\sqrt{2}/2}.$$

Le calcul a été fait en demandant seulement l'égalité

$$nv(n+1, n+2) = nv(n+1, [n])$$

(cas $\epsilon = 0$, avec les notations du § 3.2). Mais on aura une inégalité stricte en déplaçant arbitrairement peu la chaîne de points ce qui peut se faire sans modifier l'ordre de croissance de $x(n)$.

(De façon précise, il suffit de poser :

$$\forall n : dm(x) = A'(n) dm(n-1),$$

où $A'(n)$ dépasse arbitrairement peu la valeur de $A(n)$ utilisée au § 4.4.1 ; en sorte que le produit $\prod \{(A'(n)/A(n)) \mid n = 1, \dots\}$ soit arbitrairement voisin de 1).