LES CAHIERS DE L'ANALYSE DES DONNÉES

J. P. BENZÉCRI

Sur une généralisation du problème de l'ajustement d'une mesure à des marges

Les cahiers de l'analyse des données, tome 8, n° 3 (1983), p. 359-370

<http://www.numdam.org/item?id=CAD_1983__8_3_359_0>

Les Cahiers de l'Analyse des Données
Vol. VIII - 1983 - n° 3 - p. 359-370

SUR UNE GÉNÉRALISATION DU PROBLEME DE L'AJUSTEMENT D'UNE MESURE A DES MARGES

[GEN. AJUS. MARGES]

par J.P. Benzécri (1)

0 *Rappel sur l'ajustement d'une mesure à des marges* : Nous reprenons l'exposé de J.L. Madre ([*METH. AJUS. MARGES*] ; C.A.D., Vol V n° 1 ; pp 87-99 ; 1980) en suivant pour les généralisations, A. Bener ([*INTER. CORR. MULT.*] ; C.A.D., Vol VII n° 1 ; pp 25-32 ; 1982).

Considérons d'abord le cas binaire. On suppose donné deux ensembles finis J1 et J2 munis chacun d'une loi de probabilité (mesure strictement positive de masse totale 1) m_{J1}, m_{J2} ; on considère de plus sur le produit $S = J1 \times J2$, une loi q_S ; on cherche une loi p_S ayant pour marges $p_{J1} = m_{J1}$ et $p_{J2} = m_{J2}$, et de plus aussi voisine que possible de q_S. Pour préciser le problème, il faut dire en quel sens on entend "aussi voisin que possible". Une possibilité est de mesurer l'écart entre p_S et q_S par l'entropie relative $H(p_S ; q_S)$:

$$H(p_S;q_S) = \Sigma\{p_s \log_2(p_s/q_s) | s \in S\}.$$

On montre alors (Madre § 3.3) qu'il existe une mesure p_S unique réalisant le minimum de $H(p_S;q_S)$ sous la condition d'avoir des marges imposées m_{J1} et m_{J2}. De plus, cette mesure p_S peutencore être caractérisée, ses marges étant celles fixées, par la condition que sa densité (p_s/q_s) relativement à q_S, (considérée comme une fonction de $s = j1,j2$) puisse s'exprimer comme le produit d'une fonction de j1 et d'une fonction de j2. Cette dernière propriété suggère pouur la recherche de p_S un algorithme, dit RAS, sur lequel on reviendra (§§ 3 et 4).

Passons au cas ternaire : $S = J1.J2.J3$. Ici, on proposera deux généralisations du problème binaire. D' une part étant donné deux lois m_S et q_S sur S, on peut chercher une loi p_S, ayant mêmes marges simples que m_S, i.e. :

$$p_{J1} = m_{J1} \; ; \; p_{J2} = m_{J2} \; ; \; p_{J3} = m_{J3} \; ;$$

et de plus "aussi voisine que possible de q_S". D'autre part, on peut également demander que la loi inconnue p_S ait mêmes marges binaires que m_S, i.e. :

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

$$p_{J1.J2} = m_{J1.J2} \; ; \; p_{J2.J3} = m_{J2.J3} \; ; \; p_{J3.J1} = m_{J3.J1} .$$

Le lecteur aura remarqué, qu'au lieu de donner des marges, comme dans le cas binaire, nous donnons une mesure m_S et demandons que p_S ait mêmes marges que m_S : en effet (cf.[*CONTREXEMPLE MARGES*] à paraître) trois lois binaires (sur J1.J2, J2.J3, J3.J1) ne sont pas nécessairement les marges d'une loi ternaire, même si elles ont les mêmes marges simples (sur J1, J2, J3) : au contraire la donnée des marges par une mesure m_S assure, *a priori*, que le problème posé est soluble.

Dans le cas ternaire comme dans le cas binaire, si la distance entre p et q est mesurée par l'entropie relative, il y a existence et unicité de la solution ; de plus les marges de p étant imposées, cette mesure peut encore être caractérisée, par le fait que la densité (p_s/q_s) soit une fonction de s ($s = (j1,j2,j3)$) d'une certaine forme : si on impose les marges simples, la densité doit être le produit de trois fonctions dépendant chacune d'une seule des trois coordonnées j1, j2, j3 de s ; et si on impose les marges binaires, la densité doit être le produit de trois fonctions dont chacune dépend de deux des trois coordonnées j1, j2, j3. Il revient au même de dire que le logarithme de la densité peut s'exprimer comme la somme de trois fonctions (respectivement chacune d'une, ou de deux variables). Il importe de noter que ce n'est pas une condition vide : les calculs de A. Bener montrent bien qu'une fonction de s ne peut en général s'exprimer comme somme de somme de trois fonctions dépendant respectivement de (j1,j2), (j2,j3), (j3,j1).

Il est aisé de généraliser à des tableaux quaternaires ou de multiplicité supérieure à 4. L'intérêt de ces problèmes apparait dans l'analyse de correspondances multiples. Bornons-nous ici, pour la simplicité de l'exposé, au cas ternaire. On analyse communément un tableau ternaire I.J.T, en adjoignant au tableau binaire I.J, des lignes et des colonnes supplémentaires, it et jt, fournies par les étages (indicés par $t \in T$) du tableau ternaire. On obtient ainsi une représentation de l'ensemble de l'information que recèle le tableau I.J.T. Mais ici une question se pose : l'étalement sur les graphiques des points it, jt, représente-t-il proprement une interaction ternaire (au sens de [*INTER. CORR. MULT.*], ou résulte-t-il seulement de l'effet conjugué de toutes les interactions binaires; autrement dit, serait-il le même (à des fluctuations peu sensibles près) si au tableau ternaire I.J.T on substituait un tableau ayant mêmes marges binaires que celui-ci et construit d'après ces seules marges binaires. Et pour cette construction, une voie s'offre ici (explorée par C. Grossetête) : substituer au tableau donné initial appelé m_S), le tableau p_S ayant mêmes marges binaires que m_S, et de plus, aussi proche que possible de $q_S = m_I.m_J.m_T$ (loi produit des marges simples du tableau donné).

Dans la suite, on considère un "problème d'ajustement suivant des partitions" qui nous paraît comprendre tous les problèmes particuliers d'ajustement à des marges qu'on peut concevoir pour des correspondances multiples. Le "problème d'ajustement suivant des partitions" rentre lui-même dans le cadre d'un "problème linéaire général" d'ajustement.

On démontre d'abord pour le "problème linéaire général" l'existence et l'unicité d'une solution, caractérisée soit par le minimum d'une entropie relative, soit par la forme du logarithme d'une densité. Puis pour résoudre le "problème d'ajustement suivant des partitions", on propose un algorithme itératif généralisant l'algorithme RAS ; algorithme dont on démontre la convergence.

1 *Ajustement suivant des partitions, et ajustement linéaire général*

1.1 *Ajustement suivant des partitions* : On considère un ensemble fini S muni d'un ensemble EC de partitions ; une partition, élément de EC sera notée C, C' etc. . A toute mesure p_S sur S est associée une mesure induite p_C sur toute partition C de S, suivant la formule ;

$$p_C = \{p_c | c \in C\} ;$$

$$\forall \ c \in C : p_c = \Sigma\{p_s | s \in c\} .$$

A partir de EC, on définit une relation d'équivalence entre mesures sur S, qui généralise celle d'avoir certaines marges égales entre-elles ; c'est la relation d'avoir même mesure induite sur toute partition C de EC : i.e.

$$p_S \approx m_S \Leftrightarrow \forall \ C \in EC : p_C = m_C.$$

Il est clair que la relation d'avoir un certain nombre de marges en commun rentre dans ce format ; car e.g. si $S = J1.J2.J3$, la marge de p_S sur J1.J2 n'est autre que la mesure induite par p sur la partition J1.J2 de S, dont les classes sont les sous-ensembles $\{j1\}.\{j2\}.J3$.

Ceci posé, le problème d'ajustement suivant un ensemble EC de partitions s'énonce comme suit : étant donné deux lois de probabilité sur S, m_S et q_S, trouver une loi p_S telle que $\forall \ C \in EC$, $p_C = m_C$; et que sous cette condition $H(p_S ; q_S)$ soit minimum.

1.2 *Equivalence linéaire générale* : On considère un sous-espace vectoriel L^S de l'espace R^S des fonctions à valeur réelle défini sur S ; on suppose explicitement que L^S contient la droite des fonctions constantes sur S. A partie de L^S on définit une relation d'équivalence entre mesures sur S.

De façon précise étant donnée une loi m_S sur S la classe d'équivalence $L^{-1}(m_S)$ de la mesure m_S est définie par la formule :

$$L^{-1}(m_S) = \{p_S | p_S \in R_S \ ; \ \forall \ \varphi^S \in L^S \ : \varphi^S \circ p_S = \varphi^S \circ m_S\} \ ;$$

(où on a noté : $\varphi^S \circ p_S = \Sigma\{\varphi^S \ p_S \ | s \in S\}$, et de même pour $\varphi^S \circ m_S$). En d'autres termes, $L^{-1}(m_S)$ est l'ensemble des mesures sur S pour lesquelles toute fonction φ^S de L^S a même intégrale que pour la loi m_S. En imposant de plus à p_S d'être une loi sur S, on obtient ce que nous appellerons en bref la facette associée à m_S soit :

$$L^{-1}(m_S) \cap \mathfrak{L}(S),$$

(où $\mathfrak{L}(S)$ est le simplexe *fermé* des mesures sur S de masse totale 1 pour lesquelles tout point s a une masse m_s positive *ou nulle*).

1.3 *Equivalence linéaire générale et équivalence suivant une partition* : Il importe de voir que rentre dans ce format la relation introduite au § 1.1) d'avoir même mesure induite sur un ensemble EC de partition de S . Soit d'abord C une partition unique de S : toute fonction φ^C sur C , $(\varphi^C \in R^C)$, induit sur S une fonction φ^S définie comme suit :

$$\forall\ s \in S : \varphi^s = \varphi^{C(s)} ,$$

où C(s) désigne la classe de s dans la partition C. A l'espace vectoriel R^C correspond aussi un sous-espace $L^S(C)$ ensemble des fonctions sur S qui sont constantes sur chacune des classes c de la partition C . Il est clair que sont équivalentes les deux assertions suivantes:

a) m_S et p_S induisent même mesure sur C .

b) $\forall\ \varphi^S \in L^S(C) : \varphi^S \circ p_S = \varphi^S \circ m_S$.

En effet, $\varphi^S \circ p_S = \varphi^C \circ p_C$; et de même pour m .

Si maintenant on considère un ensemble EC de partitions de S , on prendra pour $L^S(EC)$ le sous-espace de R^S engendré par l'ensemble des sous-espaces $L^S(C)$ défini par une partition C , élément de EC .

En particulier, si S = J1.J2.J3, il revient au même d'imposer à une mesure p_S d'avoir mêmes marges binaires que la mesure m_S , ou d'imposer que p_S soit dans $L^{-1}(m_S)$, l'espace L^S étant ainsi défini :

L^S = sous-espaces des fonctions s'exprimant en combinaison linéaire des fonctions de deux variables (j1 et j2 ; ou j2 et j3; ou j3 et j1).

1.4 *Ajustement linéaire général* : Comme dans le cas de l'ajustement à des marges binaires (Madre § 3.2), on définit la fibre $L(q_S)$ associée à une loi q_S par une condition sur la densité :

$$L(q_S) = \{p_S | p_S \in R_S : \log(p_S/q_S) \in L^S\}$$

où par $\log(p_S/q_S)$ on note la fonction dont la valeur pour s est $\log(p_s/q_s)$.

Ceci posé, le problème général associé à L s'énonce comme suit ; étant donné deux lois de probabilités sur S , m_S et q_S , trouver une loi p_S telle que $p_S \in L^{-1}(m_S) \cap \mathfrak{L}(S)$, et que sous cette condition $H(p_S ; q_S)$ soit minimum. On montrera qu'au lieu de la condition de minimum il est équivalent de demander que $p_S \in L(q_S)$.

Aussi bien dans le calcul de $H(p_S;q_S)$ que dans la définition de $L(q_S)$, il faut prendre garde aux valeurs nulles de q_s (aux points s dont la masse pour q_S est nulle). Dans $H(p_S;q_S)$ il importe peu que p_s soit nul car $p_s \log(p_s/q_s)$ est nul ; en revanche $q_s = 0$, $p_s \neq 0$

conduit à un résultat infini. Dans la définition de $L(q_S)$ intervient la densité (p_S/q_S) dont le logarithme doit être une fonction finie de s : ce qui restreint à considérer des p_S ayant même support que q_S (i.e. ayant des normes non-nulles aux mêmes points s). Ceci conduit à poser les restrictions suivantes :

m_S étant fixé, en résulte la facette $L^{-1}(m_S) \cap \mathcal{R}(S)$: on définit le support générique de la facette comme l'ensemble S' des points s de S, où l'une au moins des mesures de la facette a une masse non-nulle ; en fait, la facette est un polyèdre convexe, et toute mesure qui est représentée par un point intérieur à ce polyèdre a pour support S'. Ceci posé on suppose désormais que q_S a pour support S' ; et en général on se restreint à considérer des mesures dont le support est S' ou est inclus dans S' : ce qui revient à restreindre S à S'.

2 *Existence et unicité de la solution pour le problème linéaire général d'ajustement* : On conserve les notations L^S, $L^{-1}(m_S)$, $L(q_S)$ posées au § 1 ; de plus, conformément à la restriction qui précède, on suppose que q_S a pour support S et que $L^{-1}(m_S)$ contient des mesures ayant pour support S ; i.e. :

$$q_S \in \mathcal{R}^+(S) \ ; \quad \mathcal{R}^+(S) \cap L^{-1}(m_S) \neq 0$$

alors on a le théorème suivant :

Le minimum $\inf\{H(p_S;q_S) \mid p_S \in L^{-1}(m_S) \cap \mathcal{R}(S)\}$ est réalisé pour une mesure unique p_S ; p_S est un point intérieur à la facette et de plus il constitue l'intersection de la fibre de q_S avec la facette : i.e.

$$p_S \in \mathcal{P}^+(S) \ ; \quad p_S = L(q_S) \cap L^{-1}(m_S).$$

Nous démontrons successivement des assertions a, b, c, en lesquelles se décompose le théorème.

a) Sur la facette le minimum de $H(p_S;q_S)$ est réalisé en un point p_S unique. En effet (cf. Madre § 3.3) sur la facette (qui est compacte), $H(p_S,q_S)$, considéré comme fonction de p_S, est une fonction de p_S, est une fonction définie continue strictement convexe ; en ce sens que que si on se restreint à faire varier p_S sur un segment inclus dans la facette ($p_S^{(t)} = t\, p_S^{(0)} + (1-t)\, p_S^{(1)}$; $t \in (0,1)$; $p_S^{(0)} \in$ facette ; $p_S^{(1)} \in$ facette) $H(p_S^{(t)} ; q_S)$ est une fonction strictement convexe de t (i.e. dont la courbe représentative est située au-dessous de chacune de ses tangentes et ne la touche qu'en un point); (voir aussi Traité TIB n° 5 § 1.3).

b) Le minimum est réalisé en un point intérieur à la facette. En effet soit $p_S^{(1)}$ un point intérieur à la facette et $p_S^{(0)}$ un point de la frontière de celle-ci (point où une masse $p_s^{(0)}$ au moins est nulle) ; alors sur le segment $(p_S^{(0)}, p_S^{(1)}$ la dérivée de $H(p_S^{(t)}, q_S)$ par rapport à t vaut $-\infty$ pour $t = 0$ (ceci résulte de la propriété correspondante de la fonction $x \log x$).

c) la différentielle de $H(p_S;q_S)$ restreinte à la facette s'annule en un point p_S intérieur à la facette $L^{-1}(m_S) \cap \mathfrak{L}(S)$ si et seulement si $p_S \in L(q_S)$.

Pour simplifier les notations nous substituons dans la définition de H le Log supérieur au $\log 2$; ceci fait, on a :

$$\partial H(p_S;q_S)/\partial p_s = 1 + \mathrm{Log}(p_s/q_s).$$

La différentielle de $H(p_S;q_S)$ pour $p_S \in R_S$, ou système des dérivées partielles calculées ci-dessus est donc :

$$\partial H(p_S;q_S)/\partial p_S = \delta^S + \mathrm{Log}(p_S/q_S) \;;$$

(où δ^S désigne la fonction constante 1). La différentielle de H restreinte à la facette s'annule en p_S si et seulement si on a pour tout vecteur tangent à la facette : i.e. $dp_S \in L^{-1}(0))$:

$$\Sigma\{(\partial H/\partial p_s) dp_s | s \in S\} = 0 \;;$$

ce qui équivaut à

$$\delta^S + \mathrm{Log}(p_S/q_S) \in L^S \;;\quad \text{où} : \mathrm{LOg}(p_S/q_S) \in L^S ,$$

(puisque $\delta^S \in L^S$ par hypothèse) ; i.e. $p_S \in L(q_S)$.

d) Si la différentielle de $H(p_S;q_S)$ (retreinte à la facette) s'annulle en p_S ; ce point p_S est le minimum absolu de H sur la facette. En effet, de la stricte convexité de H, il résulte que tout segment de la facette passant par p_S, H atteint en ce point son minimum unique.

On complétera le résultat obtenu en démontrant qu'au point p_S, la fibre $L(q_S)$ et la facette $L^{-1}(m_S)$ se coupent orthogonalement, pour la métrique du χ^2 de centre p_S ; et plus précisément que la tangente à la fibre est la sous-variété linéaire perpendiculaire en p_S à la facette.

Remarquons d'abord que la fibre $L(q_S)$ peut encore être définie comme $L(p_S)$, au sens suivant :

$$L(q_S = L(p_S) = \{f_S | f_S \in R_S \;;\; \mathrm{Log}(f_S/p_S) \in L^S\} \;;$$

si $f_S = p_S + dp_S$, la définition qui en résulte pour un vecteur dp_S tangent à la fibre est simplement :

fibre : $(dp_S/p_S) \in L^S$.

Cependant l'équation des vecteurs d'p_S tangents à la facette est :

facette : $\forall \varphi^S \in L^S : \varphi^S \circ d'p_S = \Sigma_s\{\varphi^s\ d'p_s\} = 0$

Or le produit scalaire dans la métrique du χ^2 de centre p_S est :

$<dp_S\ , d'p_S>_{pS} = \Sigma_s\{(dp_s\ d'p_s/p_s)\}$;

ce produit est nul parce que $(dp/p) \in L$; et l'espace tangent à la fibre peut être caractérisé comme l'ensemble des vecteurs perpendiculaires à tout vecteur de la facette ; cqfd.

L'orthogonalité de la facette et de la fibre, résulte encore de ce que, pour des points p'_S , q'_S voisins de l'intersection p_S l'entropie relative $H(p'_S;q'_S)$ est (à un coefficient 2 Log 2 près) équivalente au carré de la distance du χ^2 de centre p_S entre p'_S et q'_S :

$H(p'_S;q'_S) \approx (\|p'_S - q'_S\|^2_{pS}\ /(2\ \mathrm{Log}\ 2))$

3 *Algorithme RAS généralisé pour l'ajustement suivant un ensemble de partitions*

En général, il semble qu'on doive chercher le point p_S , défini par :

$\{p_S\} = L^{-1}(m_S) \cap L(q_S)$,

en se déplaçant sur la facette $L^{-1}(m_S) \cap \mathcal{P}(S)$, suivant le gradient de $N(p_S;q_S)$. Mais la méthode RAS procède en se déplaçant dans la fibre $L(q_S)$; nous montrerons que cette méthode s'étend à l'ajustement suivant un ensemble de partitions.

3.1 *Ajustement multiplicatif suivant une seule partition :* Reprenons le problème de l'ajustement d'une mesure q_S à une mesure m_S suivant un ensemble EC de partitions : au § 2 on a montré que ce problème admet une solutioon unique qu'on notera p_S. Bien que p_S soit précisément l'inconnu du problème, il est licite de se placer par rapport à p_S considérée comme origine pour étudier la convergence d'un algorithme itératif. De ce point de vue, on dira que m_S est un point de la facette de p_S ; et q_S , un point de la feuille de p_S :

$m_S \in L^{-1}(p_S) \cap \mathcal{P}(S)$; $q_S \in L(p_S) = L(q_S)$;

et l'ajustement à m_S peut aussi bien être regardé comme un ajustement à p_S.

Comme l'ajustement suivant *une seule marge*, l'ajustement *suivant une seule* partition C se réalise sans itération, par simple multiplication. Partant d'une mesure $y_S \in L(p_S) = L(q_S)$, il suffit de poser :

$C(y;p_S) = y'_S(p_C/y_C) \cdot y_S = (m_C/y_C) \cdot y_S$; i.e.

$$\forall s \in S : y'_s = (p_{C(s)}/y_{C(s)})\, y_s \quad ;$$

(où C(s) désigne, comme de régle, la classe de s dans la partition C).

Il importe de noter que l'on s'est ainsi rapproché de p_S, au sens de l'entropie relative : de façon précise on a :

$$H(p_S ; y'_S) = H(p_S ; y_S) - H(p_C ; y_C) \quad ;$$

où la quantité $H(p_C ; y_C)$ est strictement positive ; à moins que $p_C = y_C = m_C$; auquel cas l'ajustement multiplicatif suivant C était inutile et est sans effet. Il suffit d'appliquer la définition de H.

$$\begin{aligned} H(p_S ; y'_S) &= \Sigma\{p_s \log(p_s/y'_s) \mid s \in S\} \\ &= \Sigma\{\Sigma\{p_s \log(p_s/y'_s) \mid s \in S\} \mid c \in C\} \\ &= \Sigma\{\Sigma\{p_s \log((p_s/y_s)(y_s/y'_s)) \mid s \in c \mid c \in C\} \\ &= \Sigma\{p_s \log(p_s/y_s) \mid s \in S\} \\ &\quad + \Sigma\{\Sigma\{p_s \log(y_c/p_c) \mid s \in C\} \mid c \in C\} \\ &= H(p_S ; y_S) - H(p_C ; y_C). \end{aligned}$$

3.2 *Algorithme d'ajustement multiplicatif alternatif* : Posons :

$$y_S^{(0)} = q_S \; ; \; y_S^{(1)} = C^{(1)}(y_S^{(0)} ; m_S) \; ; \; y_S^{(2)} C^{(2)}(y_S^{(1)} ; m_S) ; \ldots\ldots$$

$$y_S^{(n+1)} = C^{(n+1)}(y_S^{(n)} ; m_S) \; ; \; \ldots\ldots \qquad ;$$

c'est-à-dire que, partant de $y_S^{(0)} = q_S$, on ajoute alternativement les $y^{(n)}$ à m_S (où ce qui revient au même à p_S) suivant des partitions $C^{(n+1)}$ appartenant à EC.

Certes, comme dans le cas modèle le plus simple de l'ajustement alternatif à deux marges m_{J1} et m_{J2} (cf. Madre, *op. laud.*), chaque ajustement détruit en partie l'effet du précédent (en ce sens que $C'(C(y_S , m_S) ; m_S)$ est ajusté à m_S suivant C', mais ne l'est plus suivant C) ; toutefois, comme on l'a montré, tous les ajustements s'accordent à rapprocher du but, qui est p_S.

Reste à choisir l'ordre de succession des ajustements $C^{(1)}$, $C^{(2)}$, $C^{(3)}$,... : il est vrai , (cf §3.4), que la convergence de $y_S^{(n)}$ vers p_S a lieu quel que soit cet ordre, pourvu que toutes les parties de EC interviennent une infinité de fois ; étant reprises indéfiniment dans le même ordre. Toutefois la démonstration donnée au §3.3 suppose que $C^{(n+1)}$ est choisie pour rapprocher au maximum du but ; i.e. tel que :

$$H(m_{C(n+1)} \; ; \; y^{(n)}_{C(n+1)}) = \mathrm{Sup}\{H(m_C \; ; \; y_C^{(n)}) \mid C \in EC\} \; ;$$

(la démonstration vaut encore si $H(m_{C(n+1)} ; y^{(n)}_{C(n+1)})$ représente une fraction u(n) du maximum, telle que la série de terme général u(n) diverge). On parlera alors d' "algorithme 1".

3.3 *Démonstration de la convergence de l'algorithme 1*: Il suffit de montrer que quand $n \to \infty$, tend vers zéro la distance $H(p_S ; y^{(n)}_S)$. Pour cela, on considérera simultanément une autre distance :

$$HEC(p_S ; y_S) = \Sigma\{H(p_C ; y_C) | C \in EC\}.$$

Notons k = Card EC . Il est clair qu'à chaque itération la distance H diminue d'une quantité au moins égale à HEC/(k-1) ; de façon précise on a :

$$H(p_S ; y^{(n+1)}_S) \leq H(p_S ; y^{(n)}_S) - (HEC(p_S ; y^{(n)}_S)/(k-1))$$

(en effet, HEC est une somme de k termes ; dont l'un, ajusté à l'itération n , est nul ; et dont le plus grand est celui afférent au C(n+1) choisi).

Pour établir que la suite $H(p_S ; y^{(n+1)}_S)$ tend vers 0, plus vite qu'une suite géométrique (de raison inférieure à 1), il suffit de montrer qu'il existe une constante K telle que :

$$K\, H(p_S ; y_S) \leq HEC(p_S ; y_S) ;$$

on aura alors, comme demandé :

$$H(p_S , y^{(n+1)}) \leq H(p_S , y^{(n)})(1-(K/(k-1))).$$

Reste à établir l'existence de K.

Pour cela, on remarque d'abord que seul nous intéresse pour y_S, le domaine compact Y défini par :

$$Y = \{y_S \; y_S \in L(p_S) ; H(p_S ; y_S) \leq H(p_S ; q_S)\}$$

(i.e. la partie de la fibre de p_S situé à une distance de p_S, inférieure à celle de q_S). Sur un voisinage v de l'origine p_S, les deux distances H et HEC sont équivalentes à des distances quadratiques (c'est l'équivalence classique entre distance du χ^2 et entropie relative) et donc équivalentes entre-elles ; sur Y - v, le rapport HEC/H est borné inférieurement, comme l'est tpute fonction positive ne s'annulant pas sur un compact. Ceci établit l'existence de K et donc la convergence de l'algorithme 1.

3.4 *Démonstration générale de la convergence* : Notons n = Card EC ; supposons que les partitions C éléments de EC sont numérotées de 1 à n :

$$EC = C^1 ; \ldots ; C^r ; \ldots ; C^n \;;$$

simplifions la notation $C(y_S ; p_S)$ (posée au § 3.1, pour désigner le résultat de l'ajustement de y_S à p_S suivant la partition C), en notant simplement $C(y_S)$ (étant entendu, une fois pour toutes que l'ajustement se fait à p_S ; ou, ce qui revient au même, à m_S). Alors on a un opérateur O_p :

$$O_p = C^n \circ C^{n-1} \circ \dots \circ C^1$$

qui est le composé des ajustements successifs à p_S suivant les partitions prises dans l'ordre du numérotage choisi. Notre but est de montrer que si $y_S \in L(p_S)$, la suite :

$$y_S \,;\, O_p(y_S) \,;\dots;\, O^t_p(y_S) \,;\, O^{t+1}_p(y_S) = O^t_p(y_S)) \,;\, \dots$$

converge nécessairement vers p_S. La démonstration se fait en plusieurs points :

a) L'opérateur O_p envoie la feuille $L(p_S)$ dans elle-même ; elle ne laisse fixe que le point p_S ; et rapproche strictement tout point de p_S en ce sens que :

$$y_S \in L(p_S) - \{p_S\} \Rightarrow$$

$$H(p_S \,;\, O_p(y_S)) <_s H(p_S \,;\, y_S).$$

<u>*preuve de a*</u> : Soit y_S $L(p_S)$; $C^r \in EC$: il est clair que $C^r(y_S)$ est dans la feuille $L(p_S)$; de plus (cf. § 3.1)) ou bien $C^r(y_S) = y_S$, ou bien $C^r(y_S)$ est strictement plus proche de p_S que ne l'est y_S. On voit donc que l'opérat^ur O_p qui est le composé de C^r, ne peut que rapprocher y_S de p_S ; de plus il le rapproche strictement sauf si chacune des opérations C^r successivement effectuées a laissé fixe y_S : mais cela signifie que y_S est déjà ajusté à p_S suivant toutes les partitions C^r, et par conséquent que $p_S = y_S$ (puisque la solution du problème d'ajustement est unique : cf. § 2).

b) Soit Y une partie compacte de $L(p_S)$; $p_S \notin Y$; alors il existe un nombre α strictement positif tel que :

$$\forall\, y_S \in Y \,:\, H(p_S \,;\, O_p(y_S)) \leq H(p_S \,;\, y_S) - \alpha$$

<u>*preuve de b*</u> : Sur Y la différence $H(p_S \,;\, y_S) - H(p_S ; O_p(y_S))$ est une fonction continue ; qui atteint son minimum ; lequel ne peut être nul puisque, par hypothèse $p_S \notin Y$.

c) Soit R et ε deux réels positifs ; alors il existe un entier N tel que , $\forall\, y_S \in L'p_S)$:

$$H(p_S \,;\, y_S) \leq R \Rightarrow H(p_S \,;\, O^N_p(y_S)) \leq \varepsilon$$

<u>*preuve de c*</u> : Prenons pour Y la partie compacte de $L(p_S)$ définie par :

$$Y = \{y_S \; y_S \in L(p_S) \,;\, H(p_S \,;\, y_S) \in (\varepsilon \,,\, R)\} \,;$$

soit α le nombre positif correspondant à Y dont (b) affirme l'existence : il suffit de poser :

$$N = [(R - \varepsilon)/\alpha]$$

(où [x] désigne le plus petit entier supérieur ou égal à x).

Du point (c), il résulte que la suite des $O^t_p(y_S)$ converge vers p_S (et même, que la convergence est uniforme en y_S).

Remarque : Au voisinage de p_S, l'opérateur O_p comme les C^r est tangente à une application linéaire ceci permet d'étudier explicitement la convergence : on le verra au § 4 sur l'exemple de l'ajustement à des marges.

4 *Etude élémentaire de la convergence asymptotique de l'algoritme RAS dans le cas de l'ajustement à des marges* : Pour simplifier les notations considérons le cas ternaire $S = I.J.T$; on cherche à ajuster aux trois marges binaires d'une mesure m_S donnée, la mesure initiale $q_S = m_I . m_J . m_T$, produit des marges simples de m_S. On notera y_S un point de la fibre $L(q_S)$ supposé proche du point cherché p_S. La fibre $L(q_S) = L(p_S) = L(y_S)$ est définie comme l'ensemble des mesures dont la densité par rapport à q_S (ou y_S, ou p_S) est égale au produit de trois fonctions dont chacune ne dépend que de deux des trois variables (i,j,t). Si on se place près de p_S, les densités considérées sont voisines de 1, et on peut simplifier les calculs par la formule valable au 1-er ordre : $(1+\varepsilon)(1+\varepsilon') \approx (1+\varepsilon+\varepsilon')$. De plus pour la densité $d^S = (y_S/p_S)$, on utilisera la décomposition expliquée par A. Baner en termes $pr(a)d^S$, associés aux parties a de l'ensemble à 3 éléments {I,J,T} ; mais pour abréger l'écriture on notera :

$$y_S/p_S \approx 1 + \hat{y}^I + \hat{y}^J + \hat{y}^T + \hat{y}^{IJ} + \hat{y}^{JT} + \hat{y}^{TI} ;$$

dans cette décomposition le terme constant est 1 si y_S comme p_S a masse totale 1 ; et le terme en IJT est absent parce que $y_S \in L(p_S)$.

Avec ces notations on a :

$$(y_I/m_I) = (y_I/p_I) \approx 1 + \hat{y}^I ; \ (y_J/m_J) \approx 1 + \hat{y}^J$$

$$(y_{IJ}/m_{IJ}) = (y_{IJ}/p_{IJ}) \approx 1 + \hat{y}^I + \hat{y}^J + \hat{y}^{IJ} ;$$

et des formules analogues pour les paires (J , T) et (T , I). Ceci permet de suivre (aupremier ordre) l'effet de la correction de la marge I.J de la loi y_S pour l'ajuster à celle de m_S. Si on pose :

$$y'_S = (m_{IJ}/y_{IJ})y_S ;$$

(i.e. : $y'_{ijt} = (m_{ij}/y_{ij})y_{ijt}$), on a pour la densité de y'_S :

$$(y'_S/p_S) \approx 1 + \hat{y}^T + \hat{y}^{JT} + \hat{y}^{TI} .$$

En fait, il subsiste des termes d'ordre supérieurs en i, j, et ij. Mais on conçoit qu'en ajustant alternativement aux trois marges en IJ, JT, TI, on converge rapidement vers p_S ; plus rapidement même que selon la démonstration du § 3.3. Eventuellement, on peut désirer ajuster simultanément aux trois marges binaires, pour cela il faut poser :

$$y'_S = (m_{IJ}/y_{IJ})(m_{JT}/y_{JT})(m_{TI}/y_{TI})$$
$$(y_I/m_I)(y_J/m_J)(y_T/m_T) \cdot y_S \ ;$$

où les quotients $((y_I/m_I)$ des marges simples ont été ajoutés aux quotients des marges binaires afin que les termes $\hat{y}^I$, $\hat{y}^J$, $\hat{y}^T$ ne soient pas retranchés deux fois de (y_S/p_S) (e.g. pour $\hat{y}^I$ une fois avec (m_{IJ}/y_{IJ}) et une fois avec $(m_{IT}/y_{IT}))$.

Dans le cas quaternaire et au delà, (e.g. ajustement de q_S à toutes les marges ternaires de m_S pour S = J1.J2.J3.J4), il est encore possible d'effectuer un ajustement simultané, mais la formule en est complexe et nous recommendons donc d'ajuster aux marges désirées prises une à une en reprenant celles-ci plusieurs fois de suite dans le même ordre.

Pour la pratique, il importe de noter que les ajustements successifs considérés ici restent rigoureusement dans la fibre $L(q_S)$; si donc, ce qu'il est aisé de vérifier, la convergence a lieu, c'est bien vers l'unique solution p_S intersection de $L(q_S)$ et de $L^{-1}(m_S)$.

NOTA. On a supposé au § 2 : "que q'_S a pour support S et que $L^{-1}(m_S)$ contient des mesures ayant pour support S ". Cette dernière condition peut n'être pas vérifiée dans la pratique si S = I.J.T. Soit par exemple I un ensemble de pays, J un ensemble de marchandises et T une suite d'années : k(i,j,t) = exportation de j par i pendant t . Il est vraisemblable que les marges sur I.T et J.T ne comportent pas de zéro : que chaque année t des exportations se produisent pour tout pays i ; et aussi pour toute marchandise j . En revanche, avec une nomenclature J où figurent des postes tels que céréales, bois, matériel aéronautique,... la marge I.J comportera des k(i,j) nuls : (le pays i n'exportant jamais de j). Soit donc Z' le support de la marge sur I.J :

$$Z = \{(i,j)\ i \in I\ ;\ j \in J\ ;\ k(i,j) \neq 0\}\ ;$$

il faudra se placer non dans S = I.J.T, mais dans S' = Z.T. comme loi initiale on prendra non $q_S = m_I \cdot m_J \cdot m_T$ (produit des trois marges simples du tableau ternaire donné), mais $q_{S'} = m_Z \cdot m_T$. En fait ceci revient à faire sur $q_S = m_I \cdot m_J \cdot m_T$, une itération préalable pour l'ajuster à la marge m_{IJ} ; dans la pratique la restriction du support se fait donc automatiquement !