REVUE DE STATISTIQUE APPLIQUÉE

P. CAZES

Décomposition d'un histogramme en composantes gaussiennes

Revue de statistique appliquée, tome 24, n° 1 (1976), p. 63-82

<http://www.numdam.org/item?id=RSA_1976__24_1_63_0>

DÉCOMPOSITION D'UN HISTOGRAMME EN COMPOSANTES GAUSSIENNES *

P. CAZES
Laboratoire de Statistique Université Pierre et Marie Curie (Paris VI)

I – POSITION DU PROBLEME

Le problème de la décomposition d'un histogramme en composantes gaussiennes a fait l'objet d'une abondante littérature. De nombreuses méthodes ont été proposées pour résoudre ce problème qui est d'une grande importance pratique.

Certaines de ces méthodes sont relativement complexes, comme la méthode des moments, ou la méthode du maximum de vraisemblance, ou exigent un nombre très important d'observations, comme des méthodes d'estimation stochastique. D'autres méthodes plus simples et plus opérationnelles essentiellement graphiques ont été employées avec succès quand les composantes sont bien séparées.

Dans le cas où les composantes sont mal séparées, et où le nombre d'observations est relativement faible, les méthodes précédentes ne sont pas appropriées.

Ceci nous a amené à essayer de résoudre ce problème de décomposition par deux méthodes : la première proposée par J.P. Benzécri, (cf (1)) et basée sur une série de déconvolutions successives, et la seconde proposée par Anne Schroeder, (cf (9)) et basée sur un processus itératif inspiré de la méthode des nuées dynamiques (cf (6)).

Nous exposons ces deux techniques, après avoir formulé mathématiquement ce problème de recherche de composantes gaussiennes et rappelé les principes ainsi que les limites des méthodes déjà employées.

Dans une dernière partie, nous comparons sur deux exemples les résultats trouvés avec ces deux techniques, et la méthode graphique de Bhattacharya (cf (2)).

II – FORMULATION MATHEMATIQUE DU PROBLEME

Soit Z une variable aléatoire continue de densité f (z). On suppose que la densité de Z peut se mettre sous la forme :

---

* manuscrit remis le 15.7.74, révisé en mars 1975.

Mots-clés : **composante gaussienne, histogramme, mélange**

$$\left.\begin{aligned} f(z) &= \Sigma\{ p_i \varphi (z - m_i ; \sigma_i^2) \mid i = 1 , r\} \\ \text{avec} \qquad \varphi (u , \sigma^2) &= (1/\sqrt{2\pi}\,\sigma) \exp -\{u^2/2\sigma^2\} \\ \Sigma\{ p_i \mid i &= 1 , r\} = 1 \end{aligned}\right\} \qquad (1)$$

Décomposer Z suivant ses composantes gaussiennes consiste à rechercher le système des nombres

$$(r , \{ p_i , m_i , \sigma_i \mid i = 1 , r \}) \text{ vérifiant (1).}$$

En d'autres termes, on suppose qu'on a une population formée de r sous populations, la grandeur mesurée z suivant dans la $i^{ème}$ sous population une loi normale de moyenne m_i et d'écart type σ_i, la proportion de cette $i^{ème}$ sous population dans la population totale étant égale à p_i.

On remarque que si au voisinage de z, seule la composante i a une influence non négligeable, on a :

$$f(z) \approx p_i \varphi (z - m_i ; \sigma_i^2) \qquad (2a)$$

$$t = \log f(z) \approx \log (p_i/\sqrt{2\pi}\,\sigma_i) - (z - m_i)^2/2\sigma_i^2 \qquad (2b)$$

$$u = \mathbf{d} \log \mathbf{f}\ (z)/d\ z \approx -(z - m_i)/\sigma_i^2 \qquad (2c)$$

L'équation (2b) est à la base de la méthode de H.J. Buchanan-Wollaston-Hodgeson, méthode améliorée par Tanaka (cf IV 1), tandis que la relation (2c) est à la base de la méthode de Bhattacharya (cf IV 2).

Supposons que l'on range les composantes de telle façon que

$$\sigma_1^2 \leqslant \sigma_2^2 \leqslant \ldots \leqslant \sigma_r^2 ,$$

et si * désigne le symbole du produit de convolution, on a :

$$\forall\, \sigma^2 \leqslant \sigma_1^2 : f(z) = \Sigma\{ p_i \varphi (z - m_i ; \sigma_i^2 - \sigma^2) * \varphi (z , \sigma^2) \mid i = 1 , r \} \qquad (3)$$

avec obtention d'un pic de Dirac en m_1 si $\sigma^2 = \sigma_1^2$. La formule (3) est à la base de la méthode par déconvolutions successives (cf V).

D'un point de vue pratique, on ne connait pas la densité de Z : Z n'est connu que par un N échantillon de réalisations indépendantes.

On pourra alors si on connait le nombre de composantes r, rattacher chaque observation à une classe par une procédure de reconnaissance des formes, d'où l'on déduira pour chaque classe des estimateurs de la moyenne et de l'écart type de cette classe, le nombre d'observations tombant dans cette classe permettant de calculer la proportion de cette classe dans la population totale. Cette procédure utilisée par A. Schroeder est développée en VI.

On peut également employer si le nombre d'observations est très important une procédure d'estimation stochastique où l'on fait rentrer les observations au fur et à mesure du déroulement de l'algorithme, jusqu'à ce qu'il y ait stabilisation des estimations des paramètres que l'on recherche (cf (11)).

Les deux méthodes précédentes tiennent compte de l'individualité de chaque observation, sans recours à l'histogramme empirique de Z.

On peut aussi à l'aide de l'histogramme empirique de Z, estimer f :

Si l'on découpe l'intervalle de variation de Z en n classes $I_1, I_2, \ldots, I_n$, et si q_j désigne le nombre d'observations de Z appartenant à la classe I_j, I_j étant de longueur l_j, on posera :

$$\forall\, 1 \leqslant j \leqslant n\ ; \forall\ z \in I_j : \hat{f}(z) = q_j / N\, l_j = y_j \qquad (3bis)$$

$\hat{f}(z)$ est une fois fixés $I_1, I_2, \ldots, I_n$, l'estimateur du maximum de vraisemblance de f (z), et l'on recherchera $\hat{r}$, $\{\hat{p}_i, \hat{m}_i, \hat{\sigma}_i \mid i = 1, \hat{r}\}$ tels que

$$\hat{f}(z) \approx \Sigma \{\hat{p}_i\, \varphi\, (z - \hat{m}_i\ ;\ \hat{\sigma}_i^2) \mid i = 1, \hat{r}\}$$

avec
$$\Sigma\{\hat{p}_i \mid i = 1, \hat{r}\} = 1$$

Pour ne pas alourdir les notations, et puisqu'aucune confusion n'est à craindre, nous noterons f (z) au lieu de $\hat{f}$ (z), r au lieu de $\hat{r}$ etc.

On peut alors formuler le système précédent sous une forme légèrement différente :

$$\text{trouver r,} \quad \{p_i, m_i, \sigma_i^2 \mid i = 1, r\} \quad \text{tels que :}$$

$$\forall\, 1 \leqslant j \leqslant n : f(z_j) = \frac{q_j}{N l_j} = y_j \approx \Sigma\, \{p_i\, \varphi\, (z_j - m_i\ ;\ \sigma_i^2) \mid i = 1, r\} \qquad (4)$$

z_j désignant le centre de l'intervalle I_j.

III – METHODES D'ESTIMATION CLASSIQUE

Il s'agit essentiellement de la méthode du maximum de vraisemblance, et de la méthode des moments. Pour appliquer ces méthodes, il faut supposer connu le nombre de composantes du mélange. Par ailleurs si les composantes sont de variance inégale, ces méthodes devenant fort complexes, ou même impraticables (cas de la méthode du maximum de vraisemblance, la vraisemblance présentant plusieurs singularités, correspondant au cas où les variances des différentes composantes s'annulent) n'ont pas à notre connaissance été utilisées.

Day (5) a employé ces deux méthodes dans le cas plus général d'un mélange de lois gaussiennes à p dimensions, en supposant toujours que chaque composante a même matrice variance. Le système des équations de vraisemblance est complexe et ne peut se résoudre que par une procédure itérative (par exemple une procédure de gradient réduit), même dans le cas unidimensionnel (p = 1) qui est le seul cas qui nous intéresse ici.

Dans le cas où l'on a deux composantes de même variance σ^2, de moyennes respectives m_1 et m_2, d'effectifs n_1 et n_2, Preston (8) fournit une abaque qui à partir du cœfficient de dissymétrie et d'aplatissement permet de calculer les quantités $\rho = n_1/n_2$ et $\delta = (m_2 - m_1)/\sigma$, d'ou l'on peut déduire m_1, m_2, n_1, n_2, et σ à partir des deux premiers cumulants, et donc des estimations de toutes ces quantités en remplaçant cœfficients de K. Pearson et cumulants par leurs valeurs empiriques.

Nous n'en dirons pas plus de telles méthodes, dont l'intérêt pratique est réduit, puisqu'elles ne sont bien adaptées qu'au cas où l'on a un petit nombre de composantes (2, 3 au maximum) de variance égale

IV – LES METHODES GRAPHIQUES

Ces méthodes plus ou moins empiriques et faciles à mettre en œuvre supposent que les composantes sont bien séparées, i.e. qu'au voisinage du mode d'une composante, les autres ont un effet négligeable. Elles ont l'avantage de ne pas supposer la connaissance a priori du nombre de composantes, ni de faire l'hypothèse que ces composantes ont même variance. Nous exposerons trois de ces méthodes.

IV. 1 – La méthode de Buchanan-Wollaston, Hodgeson, Tanaka (cf (3) et 10))

Les composantes étant supposées bien séparées, il y a autant de composantes que l'histogramme expérimental comporte de pics. Il suffit donc d'ajuster sur chaque pic une loi normale.

Si s_j désigne la fréquence relative des observations qui sont tombées dans la j^{eme} classe de l'histogramme ($s_j = q_j/N$ avec les notations du II), classe de centre z_j et d'amplitude l_j, et dans l'hypothèse où en z_j, seul intervient de façon non négligeable la $i^{ème}$ composante, on a d'après les relations (1) à (4) :

$$t_j = \log (s_j/l_j) = \log y_j \approx \log (p_i/\sqrt{2\pi}\,\sigma_i) - (z_j - m_i)^2/2\,\sigma_i^2$$

Au voisinage de z_j , t_j est une fonction parabolique de z_j ; il suffira donc sur la courbe donnant t en fonction de z, d'ajuster au voisinage du $i^{ème}$ mode une parabole pour obtenir les caractéristiques p_i , m_i , σ_i de la composante associée.

Après avoir ainsi ajusté graphiquement tous les modes de la courbe précédente, si la somme des pourcentages trouvés n'est pas voisine de 100, on enlève l'influence des composantes détectées, et l'on recommence le processus sur la courbe résiduelle ainsi obtenue.

IV. 2 – La méthode de Bhattacharya (cf (2))

Cette méthode est basée sur le fait déjà signalé que la dérivée logarithmique u de la densité y = f (z) d'une variable aléatoire normale Z est une fonction linéaire de z, et donc au voisinage d'un pic de l'histogramme, où seule la composante i a une influence, on a, d'après les formules (1) à (4), l'histogramme étant découpé en tranches d'égale amplitude h :

$$u_j = \frac{d \log y_j}{dz} \approx -\frac{(z_j - m_i)}{\sigma_i^2} \approx \frac{\log s_{j+1} - \log s_j}{h} = \frac{\log q_{j+1} - \log q_j}{h}$$

q_j (resp. s_j) désignant toujours le nombre des observations (resp. la fréquence relative des observations) tombant dans la $j^{ème}$ classe de l'histogramme.

On aura donc autant de composantes que le graphe de u en fonction de z comportera de portions de droite de pente négative, l'intersection de la $i^{ème}$

portion de droite négative avec l'axe des z, et la pente de cette droite permettant de définir moyenne et variance de la $i^{ème}$ composante.

Une des méthodes proposées par Bhattacharya pour estimer le pourcentage p_i de cette composante est de faire le rapport $\nu_i / N\alpha_i$, où : $-\nu_i$ est le nombre d'observations qui sont tombées dans la zone Δ_i où seule la composante i a une influence (zone qui correspond à la $i^{ème}$ portion de droite de pente négative sur le graphique précédent) $- \alpha_i = \int_{\Delta i} \varphi\,(z - m_i ; \sigma_i^2)\, dz$

N désignant toujours le nombe total d'observations, N α_i est donc le nombre d'observations qui seraient tombées dans la zone Δ i, si l'on avait eu qu'une seule composante, de moyenne m_i, et d'écart type σ_i.

Si la somme des pourcentages obtenus sur les composantes décelées est inférieure à 100, on recommence le processus après avoir retiré ces composantes.

IV. 3 – La méthode de Harding et Cassié (cf (4) et (7))

Dans cette méthode on utilise le fait que si on adopte en ordonnées une échelle galtonnienne (i.e. une échelle où la valeur de p, p appartenant à l'intervalle ouvert]0, 1[, est portée proportionnelle à u tel que

$$G(u) = p = \int_{-\infty}^{u} (1 \sqrt{2\pi}) \exp - (v^2/2)\, dv),$$

la fonction de répartition d'une loi normale de moyenne m, et d'écart type σ est une droite, la droite de Henry, la moyenne m étant l'abscisse associée à une ordonnée p de 50 %, et l'écart type σ étant tel que la distance séparant les abscisses d'ordonnées respectivement égales à 5 % et 95 % est égale à $2 \times 1{,}645\ \sigma = 3{,}29\ \sigma$.

Avec l'échelle précédente en ordonnée, on obtient pour la fonction de répartition d'un mélange de composantes gaussiennes assez séparées, une courbe qui a autant de points d'inflexion qu'il y a de composantes.

Le point d'inflexion le plus bas P_1 (i.e. d'ordonnée p_1 la plus faible) correspond à la fin de l'influence de la première composante et au début de l'influence de la seconde ; p_1 donne le pourcentage de la première composante du mélange.

Multipliant toutes les ordonnées p par $100/p_1$, pour les points situés en dessous de P_1, on obtient la droite de Henry associée à la première composante, d'où l'on déduit moyenne et écart type de cette composante.

Après avoir éliminé cette composante, on recommence le processus qui revient donc à isoler les droites de Henry associées à chaque composante.

Dans le cas où il y a une légère interférence entre deux composantes successives i et i + 1, ce dont on peut s'apercevoir, si après correction des points situés au dessous du point d'inflexion P_i associé à la composante i, on obtient une courbe différant systématiquement d'une droite, on diminue légèrement le pourcentage estimé de la composante i, et pour construire la droite de Henry associée à la composante i + 1, au lieu de partir de P_i, et de ne considérer que les points situés entre P_i et P_{i+1}, P_{i+1} désignant le

point d'inflexion situé immédiatement après P_i, on part légèrement plus bas que P_i. On voit que cette façon de procéder est assez empirique, et demande, comme le disent les auteurs du savoir faire et de l'expérience.

V – LA METHODE DES DECONVOLUTIONS SUCCESSIVES (cf (1))

V. 1 – Principe

On a vu en II que si l'on rangeait les composantes de telle sorte que $\sigma_1^2 \leqslant \sigma_2^2 \leqslant \ldots \leqslant \sigma_r^2$ la densité f (z) donnée par (1) pouvait encore s'écrire pour $\sigma^2 \leqslant \sigma_1^2$:

$$f(z) = \Sigma \{p_i \varphi (z - m_i ; \sigma_i^2 - \sigma^2) \mid i = 1, r\} * \varphi (z ; \sigma^2)$$

$$= g * \varphi$$

où g * φ désigne le produit de convolution des fonctions g et φ, et ou l'on a posé :

$$g(z) = \Sigma \{p_i \varphi (z - m_i ; \sigma_i^2 - \sigma^2) \mid i = 1, r\}$$

Pour $\sigma^2 = \sigma_1^2$, on obtient un pic en m_1, i.e. une masse de Dirac placée en ce point.

Pour $\sigma^2 > \sigma_1^2$, la déconvolution n'est plus possible.

On cherchera donc à déconvoler f par des lois normales centrées de variance σ^2 croissante, et l'on s'arrêtera quand on ne pourra plus déconvoler, et que l'on a obtenu un pic. Après retrait de ce pic, on a la fonction (en posant $\varphi (z - m_1 ; o) = \delta (z - m_1)$ masse de Dirac placée en m_1) :

$$g_1(z) = g(z) - p_1 \delta (z - m_1)$$

$$= \Sigma \{p_i \varphi (z - m_i ; \sigma_i^2 - \sigma_1^2 \mid i = 2, r\}$$

et l'on recommence le processus sur g_1 (z) que l'on peut déconvoler par des lois normales de variance $\tau^2 \leqslant \tau_2^2 = \sigma_2^2 - \sigma_1^2$; l'on obtient un pic pour $\tau^2 = \tau_2^2$ en m_2 que l'on retire, et l'on réitère le processus sur la fonction g_2 restante. On a alors ;

$$f(z) = \varphi (z, \sigma_1^2) * (p_1 \delta (z - m_1) + g_1 (z))$$

$$g_1(z) = \varphi (z, \tau_2^2) * (p_2 \delta (z - m_2) + g_2 (z))$$

$$g_2(z) = \varphi (z, \tau_3^2) * (p_3 \delta (z - m_3) + g_3 (z))$$

soit

$$f(z) = p_1 \varphi (z - m_1 ; \sigma_1^2) + p_2 \varphi (z - m_2 ; \sigma_1^2 + \tau_2^2) +$$

$$p_3 \varphi (z - m_3 ; \sigma_1^2 + \tau_2^2 + \tau_3^2) + \ldots + p_r \varphi (z - m_r ; \sigma_1^2 + \tau_2^2 + \ldots + \tau_r^2)$$

V. 2 – Réalisation de la déconvolution.

Le problème est le suivant :

connaissant les fonctions f et ψ qui sont des densités de probabilité, dans le cas qui nous intéresse, l'on désire déterminer la fonction h (qui est aussi une densité de probabilité) telle que :

$$f = \psi * h$$

soit
$$f(z) = \int_{-\infty}^{+\infty} h(\nu)\ \psi(z - \nu)\ d\nu \qquad (5)$$

h sera estimé par sa valeur en p points $\nu_1, \nu_2, \ldots \nu_p$, ce qui revient a supposer que l'on a divisé l'intervalle où h est non nul en p intervalles disjoints consécutifs $I'_1, I'_2, \ldots, I'_p$, et que dans chaque intervalle, h est constant.

Posant : $h_i = h(\nu_i)$

$$\eta_i(z) = \int_{I'_i} \psi(z - \nu)\, d\nu,$$

(5) s'écrit :

$$f(z) = \Sigma\{h_i\ \eta_i(z) \mid i = 1, p\} \qquad (6)$$

Si les fonctions f et ψ sont parfaitement connues, on déterminera les h_i de façon à ce que l'approximation de f (z) par $\Sigma\{h_i\ \eta_i(z) \mid i = 1, p\}$ soit la meilleure dans la métrique du χ^2 de centre f (z), i.e. de telle sorte que :

$$\int_{-\infty}^{+\infty} \frac{(f(z) - \Sigma\{h_i\ \eta_i(z) \mid i = 1, p\})^2}{f(z)}\, dz$$

soit minimum, en se rappelant que h étant une densité de probabilité les h_i doivent être positifs, et tels que

$$\Sigma\{h_i\ l'_i \mid i = 1, p\} = 1$$

(l'_i désignant la longueur de I'_i), ce qui revient à minimiser une forme quadratique des h_i sous les contraintes précédentes.

Dans le cas où la fonction f (z) est estimée à l'aide d'un histogramme, ce qui revient à découper l'intervalle de variation de z en n intervalles consécutifs disjoints, $I_1, I_2, \ldots, I_n$, on a d'après (5), en posant

$$F_j = \int_{I_j} f(z)\, dz$$

$$F_j = \int_{I_j} \left(\int_{-\infty}^{+\infty} h(\nu)\ \psi(z - \nu)\, d\nu \right) dz$$

$$= \Sigma\left\{ \int_{I_j} \int_{I'_i} h(\nu)\ \psi(z - \nu)\, d\nu\, dz \mid i = 1, p \right\}$$

$$= \Sigma\{h(\nu_i^*)\ \psi(z_j^* - \nu_i^*)\ l'_i\ l_j \mid i = 1, p\}$$

où (ν_i^*, z_j^*) désigne un point intérieur du rectangle $I'_i \times I_j$, l'_i (resp l_j) étant la longueur de I'_i (resp. I_j). ν_i (resp. z_j) désignant le centre de I'_i (resp. I_j),on a approximativement, puisque $h_i = h(\nu_i)$:

$$F_j \approx f(z_j)\, l_j \approx \Sigma\{ h_i\, l'_i\, \psi(z_j - \nu_i)\, l_j \mid i = 1, p\} \tag{7}$$

Posant :

$$\forall\, 1 \leqslant i \leqslant p\ ;\ 1 \leqslant j \leqslant n :$$

$$b_i = h_i\, l'_i$$

$$a_{ji} = \psi(z_j - \nu_i)\, l_j$$

(7) peut s'écrire :

$$F_j = \Sigma\{a_{ji}\, b_i \mid i = 1, p\} + e_j \tag{8}$$

e_j désignant l'erreur faire en assimilant les membres extrêmes de (7).

La détermination de b_i sera faite de telle sorte que dans R^n muni d'une métrique convenable, le vecteur $\vec{e}$ de composantes $\{ e_j \mid j = 1, n\}$ soit de norme minimum.

Choisissant la métrique du χ^2 de centre $\vec{F}$ (vecteur de composantes $\{ F_j \mid j = 1, n\}$) on est donc ramené à minimiser :

$$\Sigma\left\{ \frac{(F_j - \Sigma\{ a_{ji}\, b_i \mid i = 1, p\})^2}{F_j} \mid j = 1, n \right\}$$

sous les contraintes

$$0 \leqslant b_i \leqslant 1$$

$$\Sigma\{ b_i \mid i = 1, p\} = 1$$

Posant :

$$F'_j = \sqrt{F_j}$$

$$a'_{ji} = a_{ji}/\sqrt{F_j}$$

on est donc ramené à minimiser

$$\left.\begin{array}{l} \Sigma\{(F'_j - \Sigma\{a'_{ji}\, b_i \mid i = 1, p\})^2 \mid j = 1, n\} \\ \text{avec} \qquad \forall\ 1 \leqslant i \leqslant p : b_i \geqslant 0 \\ \Sigma\{ b_i \mid i = 1, p\} = 1 \end{array}\right\} \tag{9}$$

On résoudra (9) à l'aide d'un programme de régression sous contraintes, imposant aux coefficients de régression d'être positifs et de somme 1.

V. 3 – Réalisation pratique de l'estimation des composantes du mélange

L'histogramme dont on recherche les composantes gaussiennes, comportant n classes, le nombre p de points où l'on estime la déconvolée de cet histogramme par une loi gaussienne doit être plus petit ou égal à n.

On choisit en général n = p, ce qui permet d'avoir le même découpage pour l'histogramme et les différents histogrammes résiduels lors du processus itératif de déconvolution décrit au V 1, mais le programme mis au point permet également de choisir des valeurs de p plus petites que n. On est néanmoins assez limité dans ce dernier cas, si le nombre de composantes est assez élevé (supérieur à 3), et si le nombre de classes de l'histogramme est plus petit ou égal à 20, ce qui est le cas en général ; en effet, après avoir retiré l'influence de la première composante, l'histogramme résiduel n'a plus que $n_1 = p$ composantes ; continuant le processus de déconvolution, en imposant un découpage de la déconvolée en p_1 classes, p_1 plus petit que n_1, on obtiendra donc après détection puis élimination de la deuxième composante, un histogramme à $n_2 = p_1$ classes avec $n_2 = p_1 < n_1 = p < n$; ainsi au fur et à mesure de l'estimation des caractéristiques de chaque composante, puis de son élimination, le nombre de classes de l'histogramme résiduel diminue.

En ce qui concerne l'écart type de chaque composante, son estimation est assez aisée par le principe même de l'algorithme, puisque l'on s'arrête de déconvoler quand il y a une différence significative entre l'histogramme et l'histogramme reconstitué.

Par contre, le pic que l'on obtient sur l'histogramme déconvolé peut ne pas être très net. En outre pour éliminer l'influence de la composante décelée, on retire le pic de cet histogramme, ce qui revient à enlever de cet histogramme la partie où se trouve le pic ; ce faisant, on retire également l'influence des autres composantes sur cette partie de l'histogramme ; on surestime donc le pourcentage des premières composantes par rapport aux suivantes, et l'on diminue la précision de l'estimation des composantes, au fur et à mesure du déroulement de l'algorithme, sauf si les composantes sont bien séparées.

Remarque :

Nous avons également programmé une variante de cette méthode de déconvolution, variante où l'on essaie d'approcher f (z) par une somme pondérée de lois normales de même écart type, et de moyennes également espacées:

$$f(z) \approx f_1(z) = \Sigma \{ p_i \varphi (z - \mu_i ; \sigma^2) \mid i = 1 , p \}$$

cette approximation n'étant possible que pour $\sigma^2 \leqslant \sigma_1^2$; faisant croître σ^2, on obtiendra dans l'estimation des p_i un pic pour $\sigma^2 = \sigma_1^2$.

Après correction par déconvolution de ce pic on continuera le processus.

Cette méthode ayant donné des résultats voisins, un peu moins satisfaisants dans la détermination des pics d'ordre supérieur ou égal à 2, que la méthode exposée en détail ci-dessus, nous n'en dirons plus rien ici.

VI – LA METHODE BASEE SUR LA RECONNAISSANCE DES CONSTITUANTS DU MELANGE (cf (9))

Soit $z_1 , z_2 , \ldots , z_N$ le N échantillon de réalisations de la variable aléatoire Z que l'on sait être un mélange de r lois gaussiennes, r étant connu.

Supposons de plus que l'on sache à quelle loi gaussienne, parmi les r précédentes, appartienne chaque observation.

Si n_i ($1 \leqslant i \leqslant r$) désigne le nombre d'observations issues de la $i^{ème}$ classe, n_i/N sera une approximation du pourcentage p_i de cette composante, tandis que la moyenne et la variance empirique de cette classe seront des estimateurs (les estimateurs du maximum de vraisemblance) de la moyenne et de la variance de cette composante :

$$\hat{p}_i = n_i/N$$

$$\hat{m}_i = \bar{z}_i = \Sigma \{z_j \mid z_j \in P_i\}/n_i$$

$$\hat{\sigma}_i^2 = \Sigma \{(z_j - \bar{z}_i)^2 \mid z_j \in P_i\} /n_i$$

P_i désignant l'ensemble des observations appartenant à la $i^{ème}$ classe.

On voit que le système $\hat{\theta} = \{\hat{m}_i, \hat{\sigma}_1^2 \mid i = 1, r\}$ maximise, connaissant l'affectation de chaque observation, la vraisemblance L du N échantillon, ou son logarithme :

$$\log L = -(N/2) \log 2\pi - \Sigma \left\{ \frac{n_i}{2} \log \sigma_i^2 + \frac{1}{2} \Sigma \{(z_j - m_i)^2/\sigma_i^2 \mid z_j \in P_i\} \mid i = 1, r \right\}$$

$$= -(N/2) \log 2\pi - \Sigma \left\{ \frac{n_i}{2} (\log \sigma_i^2 + [(\bar{z}_i - m_i)^2 + \hat{\sigma}_i^2]/\sigma_i^2) \mid i = 1, r \right\}$$

Le système $\hat{\theta} = \{\hat{m}_i, \hat{\sigma}_i^2 \mid i = 1, r\}$ minimise donc

$P = (P_1, \ldots P_r)$ étant donné :

$$W(\theta, P) = -(N \log 2\pi + 2 \log L) = \Sigma \{ n_i (\log \sigma_i^2 + ((\bar{z}_i - m_i)^2 + \hat{\sigma}_i^2)/\sigma_i^2) \mid i = 1, r\}$$

avec

$$\theta = \{m_i, \sigma_i^2 \mid i = 1, r\}$$

Remarquons que $W(\theta, P)$ peut se mettre sous forme :

$$W(\theta, P) = \Sigma \{\Sigma \{(\log \sigma_i^2 + (z_j - m_i)^2/\sigma_i^2) \mid z_j \in P_i\} \mid i = 1, r\}$$

$$= -N \log 2\pi - 2 \Sigma \{ p_{ij} \log L_i(z_j) \mid i = 1, r ; j = 1, N\}$$

$$= -N \log 2\pi - 2 \Sigma \{ \log (\Sigma \{ p_{ij} L_i(z_j) \mid i = 1, r\}) \mid j = 1, N\}$$

où p_{ij} vaut 1 si z_j appartient à la $i^{ème}$ classe, 0 sinon

$$L_i(z) = \varphi(z - m_i ; \sigma_i^2)$$

étant la densité de la loi normale de moyenne m_i et de variance σ_i^2

On a vu que si P est donné, $\hat{\theta}$ minimise W ; nous poserons :

$$\hat{\theta} = g(P)$$

Supposons maintenant que l'on connaisse θ et que l'on désire affecter à chaque observation z_j une probabilité p_{ij} d'appartenance à la $i^{ème}$ classe ($1 \leqslant i \leqslant r$; $1 \leqslant j \leqslant N$).

La méthode de maximum de vraisemblance revenant à maximiser pour chaque observation j

$$\Sigma\{ p_{ij} L_i (z_j) \mid i = 1 , r\} \quad \text{avec} \quad \Sigma\{ p_{ij} \mid i = 1 , r\} = 1,$$

conduit à affecter j à la classe i (j) pour laquelle sa vraisemblance est maximale, i.e. $p_{ij} = \delta_i^{i(j)} = 1$ si i (j) = i, 0 sinon.

La partition $\hat{P}$ ainsi obtenue maximise donc, θ étant donné, parmi toutes les partitions possibles (et même parmi tous les systèmes de degré d'appartenance possible p_{ij}) la vraisemblance du N – échantillon, i.e. minimise W (θ , P).

Nous poserons

$$\hat{P} = h (\theta)$$

Des résultats précédents, l'on peut déduire un processus itératif permettant d'estimer les caractéristique $\{ m_i , \sigma_i^2 , p_i \mid i = 1 , r\}$ des r composantes gaussiennes que l'on recherche.

Se donnant une partition initiale en r classes, $\overset{\circ}{P}$ quelconque, on en déduit l'estimateur du maximum de vraisemblance $\overset{\circ}{\theta} = g (\overset{\circ}{P})$ associé à cette partition. A $\overset{\circ}{\theta}$, on peut faire correspondre la partition $\hat{P} = h (\overset{\circ}{\theta})$ qui affecte chaque observation à la classe pour laquelle sa vraisemblance est maximale.

De façon générale⊕, à l'étape n, on aura :

$$\overset{n}{P} = h (\overset{n-1}{\theta})$$

$$\overset{n}{\theta} = g (\overset{n}{P})$$

La suite $u_n = W (\overset{n}{\theta} , \overset{n}{P})$ ainsi construite est une suite finie, puisqu'on a un nombre fini N d'observations, et donc un nombre fini $(2^N - 1)$ de partitions P, et de valeurs de $\theta = g (P)$ associées, décroissante par construction, puisque :

$$u_n = W (\overset{n}{\theta} , \overset{n}{P}) = W (g (\overset{n}{P}) , \overset{n}{P}) \leqslant W (\overset{n-1}{\theta} , \overset{n}{P})$$

$$= W (\overset{n-1}{\theta} , h (\overset{n-1}{\theta})) \leqslant W (\overset{n-1}{\theta} , \overset{n-1}{P}) = u_{n-1}$$

Elle converge donc vers une valeur u* qui est atteinte pour le couple (θ* , P*) vérifiant :

$$P^* = h (\theta^*)$$

$$\theta^* = g (P^*)$$

Il faut noter que la valeur $u^* = W (\theta^* , P^*) = W (\theta^* , h (\theta^*))$ réalise un minimum local de la fonction $u (\theta) = W (\theta , h (\theta))$, minimum qui dépend donc de la partition de départ $\overset{\circ}{P}$.

On aura donc intérêt à faire plusieurs essais en choisissant des partitions initiales différentes, et en gardant le couple (θ* , P*) donnant la plus faible valeur de W (θ* , P*).

⊕ Ce processus itératif correspond au processus itératif employé en classification automatique dans la méthode des nuées dynamiques, θ jouant le rôle des noyaux.

Remarquons qu'au lieu de partir d'une partition $\overset{\circ}{P}$ pour débuter l'algorithme, on peut démarrer avec un système $\overset{\circ}{\theta}$ de caractéristiques des r lois gaussiennes, par exemple un système $\overset{\circ}{\theta}$ déduit par d'autres méthodes, comme les méthodes graphiques.

Nous désignerons par R_1 la procédure qui vient d'être exposée, pour la différencier de la procédure R_2 qui en est une variante, et qui est exposée ci-dessous dans la remarque b)

n. b.

a) L'algorithme que l'on vient de décrire est très général, car il ne fait aucune hypothèse sur la forme L de la vraisemblance du mélange étudié ; il s'applique donc à n'importe quel mélange de lois de même type, par exemple des lois gamma, ou des lois normales multidimensionnelles.

b) Au lieu d'affecter chaque individu j à une classe, on peut lui affecter une probabilité d'appartenance p_{ij} à la classe i, par exemple

$$p_{ij} = \frac{L_i(z_j)}{\Sigma\{L_i(z_j) \mid 1 = 1, r\}} ;$$

on n'optimise plus, θ étant fixé, $W(\theta, P)$ où P désigne ici le système des $\{p_{ii} \mid 1 \leqslant i \leqslant r ; 1 \leqslant j \leqslant N\}$, mais l'on peut par ce procédé reconnaître des populations très imbriquées, et obtenir un meilleur ajustement de l'histogramme expérimental des z_j.

Dans ce cas, pour passer de P à θ, on posera :

$$p_i = \Sigma\{p_{ij} \mid j = 1, N\} \; ; \; m_i = (1/p_i)\,\Sigma\{p_{ij}\, z_j \mid j = 1, N\}$$

$$\sigma_i^2 = (1/p_i)\,\Sigma\{p_{ij}\,(z_j - m_i)^2 \mid j = 1, N\}$$

Nous désignerons par R_2 cette procédure.

VII – EXEMPLES D'APPLICATION

VII. 1 – Premier exemple

Cet exemple a été emprunté à Bhattacharya (cf (2)) : il s'agit d'un mélange artificiel de trois lois gaussiennes dont on a l'histogramme qui comporte 18 classes (cf tableau 1).

On a respectivement appliqué la méthode de Bhattacharya, la méthode des déconvolutions successives, et la méthode du VI d'A. Schroeder (procédure R_1). Dans ce dernier cas, on a simulé l'histogramme qui comporte 12.927 individus en faisant l'hypothèse que dans chaque classe de l'histogramme, les observations ont une distribution uniforme. On a également fait des essais d'une part en ne simulant l'histogramme qu'avec 6.460 individus, dans le cas où l'on applique le processus itératif du VI, en affectant des probabilités d'appartenance (procédure R_2), d'autre part en divisant l'histo-

gramme en 9 et 6 classes respectivement, i.e. en prenant un pas respectivement double ou triple du pas initial.

Tableau 1

Histogramme d'un mélange artificiel de 3 populations gaussiennes

n°	classe	centre	fréquence observée N_j
1	8- 9	8,5	31
2	9-10	9,5	532
3	10-11	10,5	2 198
4	11-12	11,5	2 297
5	12-13	12,5	685
6	13-14	13,5	494
7	14-15	14,5	1 188
8	15-16	15,5	1 479
9	16-17	16,5	938
10	17-18	17,5	486
11	18-19	18,5	537
12	19-20	19,5	702
13	20-21	20,5	664
14	21-22	21,5	431
15	22-23	22,5	192
16	23-24	23,5	59
17	24-25	24,5	12
18	25-26	25,5	2
			N = 12 927

Dans ce dernier cas la méthode de Bhattacharya ne fournit qu'une seule composante, tandis que la méthode par déconvolutions successives s'applique mal, le nombre de classes de l'histogramme étant trop petit pour que l'on obtienne des pics nets et facilement estimables.

Ce n'est que dans ce dernier cas (histogramme à 6 classes, simulé sur 6.460 individus) que le processus R_2 du VI avec affectation de probabilités s'est stabilisé en moins de 50 itérations, aussi nous ne donnerons les résultats relatifs à cette façon d'opérer que dans ce cas là.

Tableau 2

Résultats obtenus dans l'exemple 1 (histogramme à 18 et 9 classes)

méthode	valeurs Théoriques	Bhattacharya	Deconvolutions (V)	procédure R_1 du VI	procédure R_1 du VI	Battacharya	Deconvolutions (V)	procedure R_1 du VI	procedure R_1 du VI
nB de classes de l'histogramme : n		18	18	18	18	9	9	9	9
nB d'observations N		12 927	12 927	12 927	6 460	12 927	12 927	12 927	6 460
nB d'itérations				16	14			16	13
T	69,3	40,9	297	75,8	37,9*	453	52,5	540	244*
m_1	11,04	11,03	11,04	11,04	11,04	11,22	11,10	10,96	10,96
σ_1	0,78	0,80	0,74	0,90	0,90	0,92	1,01	0,98	0,98
p_1	43,11	44,04	43,34	44,61	44,52	42,20	45,88	43,61	43,58
m_2	15,31	15,28	15,27	15,27	15,27	15,13	15,49	14,90	14,94
σ_2	1,16	1,12	1,18	1,03	1,04	1,56	1,31	0,90	0,94
p_2	34,44	33,34	35,86	32,51	32,76	37,62	34,31	28,54	29,38
m_3	19,85	19,86	19,86	19,90	19,30	19,82	20,08	19,44	19,53
σ_3	1,60	1,60	1,27	1,54	1,52	1,62	1,50	1,83	1,81
p_3	22,45	22,62	20,80	22,88	22,72	20,18	19,81	27,85	27,04

* valeurs à multiplier par $\frac{12\,927}{6\,460} = 2{,}00$ pour les comparer avec les précédentes.

R_1 : procédure du VI (sans affectation de probabilités)
R_2 : " " (avec " " ")

Tableau 3

Résultats obtenus dans l'exemple 1 (histogramme à 6 composantes).

méthode	valeurs théoriques	procédure R_1 du VI	procédure R_1 du VI	procédure R_2 du VI
nB de classes de l'histogramme n		6	6	6
nB d'observations N		12 927	6 460	6 460
nB d'itération		16	14	43
T	69,3	832	393*	144*
m_1	11,04	10,66	10,54	10,39
σ_1	0,78	1,46	1,40	1,46
p_1	43,11	39,61	37,51	32,75
m_2	15,31	15,07	14,93	14,36
σ_2	1,16	1,16	1,21	1,57
p_2	34,44	37,29	38,94	34,10
m_3	19,85	19,97	19,95	18,59
σ_3	1,60	1,80	1,81	2,69
p_3	22,45	23,10	23,54	33,15

* valeurs à multiplier par $\frac{12\,927}{6\,460} = 2{,}00$ pour les comparer avec les précédentes.

R_1 : procedure de VI (sans affectation de probabilités)

R_2 : " " " (avec " " ")

Les tableaux 2 et 3 résument les résultats obtenus.

Les composantes étant bien séparées, on a obtenu par la méthode du VI les mêmes résultats que l'on parte de partitions initiales différentes, ou que l'on parte des valeurs théoriques ou des valeurs trouvées par Bhattacharya pour le système θ des $\{m_i , \sigma_i^2 \mid i = 1 , r\}$

Pour comparer la qualité des estimations obtenues, on a calculé le χ^2 d'ajustement :

$$T = \Sigma \{(N_j - \hat{N}_j)^2 / N_j \mid j = 1 , n\}$$

où $\hat{N}_j$ est l'approximation de N_j réalisée en remplaçant l'histogramme expérimental par le mélange gaussien estimé, et n le nombre de classes de l'histogramme

$$\hat{N}_j = N \Sigma \{p_i \varphi (z_j - m_i \ ; \sigma_i^2) \mid i = 1 , r\}$$

z_j désignant le centre de la $j^{ème}$ classe de l'histogramme expérimental.

Dans le cas de l'histogramme à 18 classes, l'on voit que le meilleur ajustement est réalisé par la méthode de Bhattacharya, ce qui est normal, puisque les composantes étant bien séparées, l'ajustement graphique est meilleur ; par contre avec l'histogramme à 9 classes, la méthode par déconvolutions est plus précise.

En employant la procédure R_2 du VI, on n'a pas obtenu, comme on l'a déja dit plus haut, la convergence en 50 itérations, dans le cas d'un histogramme à 18 ou 9 classes, et l'ajustement obtenu après ces 50 itérations était moins bon que dans le cas où l'on n'affecte pas de probabilités.

Par contre dans le cas de l'histogramme à 6 classes, la procédure R_2 a convergé, et fourni un ajustement meilleur que celui obtenu sans affectation de probabilités (cf tableau 3).

Remarque :

Nous avons obtenus des résultats et des conclusions similaires avec un second lot de données (histogramme à 21 classes) fourni par Bhattacharya, et où il y avait 5 composantes gaussiennes bien séparées.

VII. 2 – Second exemple

On a simulé sur 200 observations, un mélange de deux lois gaussiennes de caractéristiques

$$m_1 = \sigma_1 = 1$$

$$m_2 = 1{,}5 \quad ; \sigma_2 = 2{,}5$$

$$p_1 = p_2 = 0{,}5$$

La méthode de Bhattacharya (avec un histogramme à 10 classes) ne permet pas de déceler les deux composantes.

La méthode par déconvolutions successives (toujours avec un histogramme à 10 classes) trouve bien le premier pic, mais elle surestime son importance, et en fait on trouve deux autres pics.

C'est la méthode du VI (procédures R_1 et R_2) qui a fourni les meilleurs résultats. Quand on demande deux composantes, la procédure R_1 fournit plusieurs résultats suivant le point de départ choisi. Par contre, la procédure R_2 a toujours convergé vers le méme résultat, et l'ajustement obtenu est meilleur que ceux obtenus avec R_1.

Si l'on impose 3 classes, que l'on parte de la partition P obtenue à partir des observations rangées par valeurs croissantes et divisées en trois

classes d'égal effectif (valeurs faibles – valeurs moyennes – valeurs fortes) ou des résultats obtenus à partir de P avec la procédure R_1, la procédure R_2 donne le même résultat.

Par contre, la procédure R_1 donne des résultats différents que l'on parte de P, ou des résultats issus de R_2, la valeur du critère u* (cf VI) étant sensiblement la même dans les deux cas, l'ajustement étant meilleur lorsqu'on part de P.

Les résultats obtenus avec ces deux procédures ne donnent en fait que 2 classes, la troisième ayant un pourcentage négligeable (1 %) ; l'ajustement et les résultats obtenus sont encore les meilleurs quand on applique la procédure R_2.

Signalons qu'en partant d'un système de probabilités d'appartenance p_{ij}, on a obtenu, au bout de 50 itérations, avec la procédure R_2 (qui n'a pas convergé dans ce cas), un meilleur ajustement que ceux précédemment obtenus, alors qu'en fait les résultats obtenus sont moins bons, les trois composantes ainsi estimées étant en proportions sensiblement égales.

Les différents résultats obtenus sont résumés sur le tableau 4.

VII. 3 – Quelques remarques pratiques et méthodologiques

Le programme utilisant la méthode de Bhattacharya est très rapide, mais il semble plus précis encore d'opérer graphiquement pour mieux intégrer visuellement les zones où l'on a des points répartis sur une droite de pente négative (cf IV 2).

En ce qui concerne le programme basé sur la méthode des deconvolutions, il semble qu'il faille prendre comme critère d'arrêt des déconvolutions, non pas un seuil pour le χ^2 d'ajustement (i.e. pour la quantité T introduite en VII 1), mais le moment où ce χ^2 au lieu de croître regulièrement, fait un saut important. En effet, si l'on a un très grand nombre N d'observations, ce χ^2 (qui est homogène à N) sera très élevé, et très vite significatif, alors que s'il y a peu d'observations, ce χ^2 sera faible.

Par ailleurs si l'on n'a pas un pic très net, la recherche automatique du pic est délicate, et il sera nécessaire de vérifier le pic adopté, et le cas échéant de repasser le programme en adoptant un autre pic ; la procédure n'est donc pas complétement automatique. Dans ce cas l'estimation de la composante gaussienne associée est douteuse ; ceci se produit pour les pics d'ordre supérieur, et il vaudra mieux alors se contenter des premières composantes déjà estimées, même si l'histogramme résiduel ne se réduit pas à zéro.

En ce qui concerne les procédures R_1 et R_2, elles ont été programmées en un seul sous programme (la procédure R_1 étant un cas particulier de la procédure R_2). Si la procédure R_2 ne se stabilise pas, on choisira comme estimation celle qui fournit le meilleur ajustement, ce qui se produit en général dès les premières itérations du processus.

Du point de vue méthodologique, on pourra opérer comme suit :

1) On applique la procédure (sans affectation de probabilités), i.e. la procédure R_1, en démarrant sur la partition P obtenue à partir des individus

Tableau 4

Exemple 2 : mélange de 2 lois gaussiennes, simulé sur 200 observations.

Essai	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Méthode	valeurs ⊕ théoriques	Déconvolutions (V)	R_1	R_1	R_2	R_2	R_2	R_1	R_1
u*			121	124,3				109,9	110,1
nB d'iterations			8	18	24 ou 23	50	32 ou 31	11	15
point de départ			P	⑤⊗	P ou ③	degré d'appartenance	P ou ⑨	⑦	P
T		9,66	27,35	36,57	5,59	3,75	4,92	34,05	24,95
m_1	1 (0,97)	1,086	0,26	0,72	1	0,84	0,88	0,70	0,37
σ_1	1 (0,9)	1,073	1,19	0,46	0,93	0,84	0,96	0,45	0,85
p_1	50	84,57	59,50	45,50	51,40	33,60	50,60	44,50	58,50
m_2	1,5 (1,83)	4	3,09	1,97	1,83	1,42	2,10	2,11	3,09
σ_2	2,5 (2,5)	1,20	1,45	2,40	2,49	1,32	2,14	2,13	1,44
p_2	50	12,59	40,50	54,50	48,60	33,80	48,40	54,50	40,50
m_3		6,9				1,96	−6,16	−6,16	−6,16
σ_3		1,56				2,82	0,8	0,8	0,8
p_3		2,84				33,60	1	1	1

⊕ valeurs simulées entre parenthèses

R_1 : procédure du VI (sans degré d'appartenance)

R_2 : (avec " ")

P : partition obtenue en divisant en classes d'égal effectif les observations rangées par valeurs croissantes.

⊗ : ⑤ = départ sur les résultats obtenus dans l'essai n° (5)

rangés par valeurs croissantes, puis divisés en r classes de même effectif, si l'on recherche r composantes.

2) On applique la procédure (avec affectation de probabilités), i.e. la procédure R_2, en partant du résultat du 1)

3) On applique cette même procédure R_2 en partant de la partition P

4) On applique la procédure R_1, en partant du résultat du 3.

L'on retient alors les résultats correspondant au meilleur ajustement.

Il faut néanmoins remarquer que cette façon de procéder ne conduit pas forcément à la meilleure solution, en particulier si le nombre r de composantes choisi, est plus fort que le nombre réel.

Pour essayer de déterminer r, on peut d'abord faire $r = 1$, et effectuer un test classique de normalité. Si ce test est significatif, on entame la procédure avec $r = 2$, puis $r = 3, \ldots$ etc, et l'on s'arrêtera quand le χ^2 d'ajustement sera inférieur à un seuil donné.

VII. 4 – Conclusion

Des deux exemples que nous avons exposés, et d'un grand nombre d'essais que nous avons effectués, il semble ressortir la philosophie suivante :

Si les composantes sont bien séparées, visibles à l'œil nu sur l'histogramme, les méthodes graphiques semblent les plus naturelles et les meilleures, car elles fournissent un excellent ajustement.

Si les composantes sont bien imbriquées, la méthode par déconvolutions successives permet d'estimer avec précision le premier pic (moyenne et écart type) ; par contre, comme elle surestime son importance, la recherche des pics suivants est délicate, et leur estimation est mauvaise*.

C'est la méthode de reconnaissance des formes du VI (combinaison des procédures R_1 et R_2) qui semble la plus adaptée, quand les composantes sont bien imbriquées, et qu'on connait leur nombre.

BIBLIOGRAPHIE

[1] BENZECRI J.P. – La régression (publication du Laboratoire de Statistique Mathématique) 1970

[2] BHATTACHARYA – A simple method of resolution of a distribution into Gaussian components (pp 115-135) *Biometrics,* Mars 67

[3] BUCHANAN-WOLLASTON H.G. et HODGESON W.G. – A new method of treating frequency curves in fishery statistics, with some résults. *j. Cons.* 4, (pp 207-225), 1929

---

* En fait cette méthode semblerait plus applicable, si on avait des histogrammes ayant 40 ou 50 classes, auquel cas on aurait probablement des pics plus nets, mais il est rare qu'on ait assez d'observations pour obtenir un histogramme à 40 ou 50 classes précis ; même si c'était le cas, le temps de calcul deviendrait prohibitif, puisqu'á chaque déconvolution, on ferait une régression sous contrainte avec 40 ou 50 coefficients à estimer.

[4] CASSIE R.M. – Some uses of probability paper for the graphical analysis of polymodal frequency distributions. *Aust. j. Mar. Freshw. Res.* 5, (pp 513-522). 1954

[5] DAY N.E. – Estimating the components of a mixture of normal distributions. *Biometrika* 56.3 (pp 463-474) 1969

[6] DIDAY E. – Une nouvelle méthode en classification automatique et reconnaissance des formes : la méthode des nuées dynamiques ; *R S A*, vol XIX, n° 2, 1971

[7] HARDING J.F. – The use of probability paper for the graphical analysis of polymodal frequency distributions. *J. Mar. biol. Ass.* U.K. 28 (pp 141-153). 1949

[8] PRESTON E.J. – A graphical method for analysis of statistical distributions into normal components. *Biometrika* 40, (pp 460-464). 1949

[9] SCHROEDER A. – Reconnaissance des composants d'un mélange. Thèse de 3° cycle, Paris VI, Mai 1974

[10] TANAKA S. – A method of analysing a polymodal distribution and its application to the lenght distribution of the Porgy, Taius tumifroms (J. et S.). *J. Fish. Res. Bd Can.* 19, (pp 1143-1159) 1962

[11] YOUG T.Y. et CORALUPPI G. – Stochastic Estimation of a mixture of normal density functions using an information criterion. *IEEE Trans on Information Theory,* Vol II-16 n° 3 (pp 258-263). Mai 1970