ESTIMATEUR D'UN MÉLANGE (C8, H2)
(15 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

Comme dans le cas de l'estimation d'une **densité** (cf **estimateur de la densité**), l'**estimation** d'un **mélange de lois** peut être paramétrique ou non (cf **estimation non paramétrique**). Elle s'effectue généralement à partir des **densités** du mélange, ou parfois à partir de ses **fonctions de répartition**.

Si $\mathcal{X}$ est un ensemble d'observation (cf **espace d'observation**) et si le mélange concerne un nombre fini de lois, la **densité du mélange** est de la forme (mélange fini) :

(1) $\quad f(x) = \Sigma_{i \in I} \alpha_i . f_i(x), \qquad \forall x \in \mathcal{X},$

où I est un ensemble d'**indices** fini, f_i est la densité de la i-ième composante du mélange et $\alpha = (\alpha_i)_{i \in I} \in S_{|I|}$ (**simplexe** de $\mathbf{R}^{|I|}$), avec |I| = Card I.

Dans le cas **paramétrique**, chaque densité f_i peut posséder un **paramètre**, eg $\theta_i \in \Theta_i$, où $\Theta_i \subset \mathbf{R}^{Q(i)}$ et Q(i) désigne Q_i (dimension du paramètre θ_i , $\forall$ i = 1 ,... , Q_i . Le paramètre d'ensemble est alors $\gamma = (\theta, \alpha)$, avec $\theta = (\theta_i)_{i \in I}$ et $\alpha \in S_{|I|}$.

Le problème principal tient, le plus souvent, au **défaut d'identifiabilité** des paramètres de f, même lorsque f est elle-même identifiable (cf **modèle identifiable**).

(i) Dans le cas non paramétrique (ou semi-paramétrique, puisque f est partiellement indexé par α), il est possible de transposer les méthodes classiques d'estimation d'une densité (**méthode du noyau**, **méthode des fonctions orthogonales**, etc) à l'estimation de la densité d'un mélange.

Ainsi, la **méthode du noyau** conduit à estimer f à l'aide d'un **estimateur** de la forme :

(2) $\quad \tilde{f_N}(x) = \Sigma_{i \in I} \tilde{\alpha_i} . (N h_{N(i)})^{-1} . \Sigma_{n=1}^{N} p_i \{(x - X_n) / h_{N(i)}\}, \qquad \forall x \in \mathcal{X},$

où $X = (X_1 ,..., X_N)$ est un N-**échantillon iid** dont les coordonnées X_n sont chacune tirée selon la loi du mélange (dont la densité est f), et où N(i) désigne N_i .

(ii) Dans le cas paramétrique, les densités composantes f_i ($i \in I$) contiennent chacune un paramètre propre θ_i et le mélange légal f peut s'écrire :

(3) $\quad f(x, \gamma) = \Sigma_{i \in I} \alpha_i . f_i(x, \theta_i), \qquad \forall x \in \mathcal{X},$

avec $\gamma = (\alpha, \theta)$, $\alpha = (\alpha_i)_{i \in I}$, $\theta_i \in \Theta_i$ (avec $\Sigma_{i \in I} Q_i = Q$) et $\theta = (\theta_i)_{i \in I}$.

Les méthodes d'estimation usuelles (**méthode du maximum de vraisemblance** ou **méthode des moments**) doivent être ici contraintes, car :

(4) $\gamma \in \Gamma = \{(\alpha, \theta) : \alpha \in S_{|I|}\ ,\ \theta \in \mathbf{R}^Q\}$,

en supposant que $\theta_i \in \mathbf{R}^{Q(i)}$ et que $\mathbf{R}^Q = \Pi_{i \in I}\ \mathbf{R}^{Q(i)}$ (avec $\Sigma_{i \in I}\ Q_i = Q$), où Q(i) désigne $Q_i \in \mathbf{N}^*$.

Par suite, la **vraisemblance** du modèle mélangé est de la forme :

(5) $L(X, \gamma) = \Pi_{n=1}^{N} f(X_n\ ,\ \gamma)$.

La **méthode du maximum de vraisemblance contraint** consiste alors à maximiser L (X, γ) pr à γ sur l'ensemble des contraintes Γ.

Souvent, en pratique, des **singularités** apparaissent dans les calculs numériques.

(iii) Dans le cadre de la **théorie bayésienne**, on dote l'ensemble $\mathbf{R}^{|I|}$ x $\mathbf{R}^Q$ d'une **probabilité a priori** qui charge seulement Γ (eg une probabilité dont la **loi marginale** en α est une **loi de DIRICHLET**). On maximise alors pr à (α, θ) la probabilité a posteriori qui en résulte.

L'équation (3) peut s'interpréter comme suit :

(a) x est la valeur d'une va qui suit a priori la loi définie par f ;

(b) la probabilité que x provienne de la i-ième « population » (ie de la i-ième loi), de densité f_i , n'est autre que α_i (cf **classification**, **analyse discriminante**).