Sur certaines lois de probabilité

par G. DARMOIS (réinterprété)

(cf **problème du NIL**)

(i) Soit $P_\theta^{(\xi, \eta)}$ la **loi de probabilité** d'un **couple aléatoire** (ξ, η) dépendant d'un **paramètre** θ, et dont la **densité** pr à λ_2 est notée h (x, y, θ).

Quel que soit θ, la **probabilité** du plan $\mathbf{R}^2$ est égale à l'unité mais, pour toute région $R \subset \mathbf{R}^2$, la probabilité $P_\theta^{(\xi, \eta)}$ (R) varie, en général, avec θ.

Ainsi (R.A. FISHER), la **loi** suivante :

(1) $\quad h(x, y, \theta) = \exp\{-\theta x - (y / \theta)\} \,.\, d(x, y),$

dont le **support** est le premier quadrant $\mathbf{R}_+^2$, est tq toute bande (**partie**) infinie R limitée par deux hyperboles {x . y = constante} possède une probabilité indépendante de θ.

(ii) On appelle **problème du Nil** (R.A. FISHER) le problème associé à la division de tout le « champ » (espace) en « parcelles » (parties) ayant cette propriété, ie dont la masse totale reste la même, quelles que soient les variations du paramètre.

(iii) Une telle fonction n'existe généralement pas. Lorsqu'elle existe, la solution de ce problème est une fonction de (x, y) dont la loi ne dépend pas de θ, et sa forme générale est du type (cf **conditionnement**, **loi conditionnelle**) :

(2) $\quad f(X) \,.\, g_X(Y, \theta).$

Le premier facteur (densité de X) ne dépend pas de θ ; ce paramètre ne figure que dans la loi conditionnelle de Y. Un changement de variables résoud le problème.

(iv) Des solutions particulières intéressantes peuvent être obtenues comme suit. Pour qu'une « parcelle » soit de probabilité invariante (pr à θ), il faut et il suffit qu'à chaque valeur de θ soit associée une transformation ponctuelle qui échange les points de la parcelle en conservant l'élément de probabilité.

En particulier, une translation :

(3) $\quad x_1 = x_2 , \qquad y_1 - \theta_1 = y_2 - \theta_2 ,$

peut s'associer à des lois dont la densité est de la forme :

(4) $\quad d(x, y - \theta).$

La loi (1) se ramène à la forme « canonique » (4) par le changement de « variables » $(x, y, \theta) \mapsto (X, Y, \Theta)$:

(5) $x\, y = e^X, \quad y / x = e^Y, \quad \theta = e^{(\Theta / 2)}$.

Ainsi, l'exemple, associé à une rotation, de la **loi circulaire** gaussienne (cf **loi gaussienne**) chargeant un cercle de centre inconnu et de rayon connu r, conduit à la forme :

(6) $c \,.\, \exp \{- (1/2) ((x - r \,.\, \cos \theta)^2 + (y - r \sin \theta)^2)\}$,

dans laquelle $c \in \mathbf{R}_+^*$ désigne une **constante**.

(v) De même, les densités des lois de la forme (2) peuvent être associées aux densités de la forme :

(7) $f (X, \theta) \,.\, g_X (Y)$,

dans lesquelles le paramètre θ ne figure plus dans la **densité conditionnelle** de Y, et où la va X donne toute l'**information** sur le paramètre.

Ces lois ont été appelées (G. DARMOIS) ***« lois à estimation exhaustive »***.

(vi) Disposant de N observations $(X_n , Y_n)_{n=1,\dots,N}$, indépendantes entre elles, du couple (ξ, η), on peut exhiber une **statistique** (ie une fonction des observations) qui ne laisse échapper aucune information relative à θ (cf **exhaustivité**). La forme générale de la solution, ie :

(8) $\exp \{\alpha (x, y) \,.\, a (\theta) + \beta (x, y) \,.\, b (\theta) + \gamma (x, y) \,.\, c (\theta)\}$,

admet pour **statistique exhaustive** le couple :

(9) $A_1 = \Sigma_{n=1}^N \alpha (X_n , Y_n), \quad A_2 = \Sigma_{n=1}^N \beta (X_n , Y_n)$.

La loi (1) est de ce type avec pour statistique exhaustive (couple constitué des totaux empiriques) :

(10) $A_1 = \Sigma_{n=1}^N X_n , \quad A_2 = \Sigma_{n=1}^N Y_n$,

et la loi du couple (X, Y), avec $X = (X_n)_{n=1,\dots,N}$ et $Y = (Y_n)_{n=1,\dots,N}$:

(11) $\{(n - 1) !\}^{-2} \,.\, x^{n-1}\, y^{n-1} \,.\, \exp \{- \theta\, x - (y / \theta)\}$,

qui joue le rôle du premier facteur dans (5), fournit toute l'information relative à θ.

(vii) La loi associée à (11) est également une solution du problème du Nil. Elle prend la forme canonique par le changement de variables précédent, avec :

(12) $\alpha = \xi . \eta, \quad \beta = \eta / \xi.$

Le paramètre θ ne figure que dans la loi conditionnelle de β.

Les simplifications qui en résultent dans le **problème d'estimation** de θ peuvent se retrouver dans d'autres solutions. On appellera :

(a) solution de type (1) une solution du problème du Nil associée à une statistique exhaustive ;

(b) solution de type (2) une solution (1) dont la statistique exhaustive est, elle-même, une solution du problème du Nil.

On peut trouver toutes les solutions de type (1) de la forme :

(13) $e^{\varphi (x , y - \theta)}$.

Deux cas, simples mais assez généraux, sont les suivants :

$(b)_1$ φ est un polynôme entier quelconque, eg de la forme :

(14) $\varphi = - \{x^4 + x^2 (y - \theta)^2 + (y - \theta)^4\}$;

$(b)_2$ φ est de la forme :

(15) $\varphi = \Sigma_i A_i (x) . \exp \{\alpha_i (y - \theta)\}$.

La solution de type (1) est de cette deuxième forme.

Par ailleurs, la loi de Gauss (6) est une solution de type (2), car elle admet une statistique exhaustive (couple des **moyennes empiriques**) :

$$(16) \quad \begin{cases} M_X = \bar{X}_N = N^{-1} \Sigma_{n=1}^{N} X_n , \\ M_Y = \bar{Y}_N = N^{-1} \Sigma_{n=1}^{N} Y_n , \end{cases}$$

dont la **loi jointe** possède une densité de la forme :

(6)' $c . \exp \{- (n / 2) ((x - r . \cos \theta)^2 + (y - r \sin \theta)^2)\}$,

dans laquelle $c \in \mathbf{R}_+^*$ désigne une constante.

Ainsi, comme pour la loi correspondant à (6), les régions à masse probabiliste invariante sont des couronnes circulaires, et le paramètre θ ne figure que dans la loi conditionnelle de l'angle polaire du point (X, Y).

(viii) Le problème du Nil se relie à la **théorie de l'estimation** et à la **théorie des tests** (qui permet de juger si la loi ainsi estimée est en accord satisfaisant avec les observations).

La théorie de J. NEYMAN - E.S. PEARSON se développe d'abord avec des lois à estimation exhaustive, puis avec les lois des types précédents : les régions R à masse invariante sont dénommées ***« régions semblables à l'espace entier »***.