REVUE DE STATISTIQUE APPLIQUÉE

JEAN-FRANÇOIS PHELIZON

# Axe principal de trois variables liées

*Revue de statistique appliquée*, tome 24, nº 1 (1976), p. 83-86

<http://www.numdam.org/item?id=RSA_1976__24_1_83_0>

# AXE PRINCIPAL DE TROIS VARIABLES LIÉES

Jean-François PHELIZON

*On se propose ici de déterminer les paramètres du premier axe principal (souvent appelé droite de régression orthogonale) d'un nuage de points dont les coordonnées sont liées par une relation linéaire. La solution, remarquablement simple dans ce cas, ne nécessite pas de devoir programmer un algorithme classique pour obtenir le vecteur propre associé à la plus grande valeur propre de la matrice des covariances, vecteur qui définit le premier axe principal du nuage des points. On montre ensuite que l'équation obtenue (celle d'une droite dans un espace à trois dimensions), caractérise une droite dans un* diagramme triangulaire, *ce qui facilite beaucoup l'interprétation du calcul.*

## I – PARAMETRES DE L'AXE PRINCIPAL.

Soient n points dans un espace à trois dimensions ; ils appartiennent au plan $\Pi_1$ d'équation

$$x + y + z = k \ (*) \tag{1}$$

On cherche le premier axe principal du nuage des points. Cet axe est évidemment contenu tout entier dans $\Pi_1$. Il appartient aussi à un plan $\Pi_2$ perpendiculaire à $\Pi_1$ et d'équation

$$ax + by + cz = d,$$

équation où l'on choisit a, b, c en sorte que

$$a^2 + b^2 + c^2 = 1.$$

Puisque $\Pi_1$ est perpendiculaire à $\Pi_2$, on a d'autre part

$$a + b + c = 0.$$

---

(1) Article remis en Avril 1974, révisé en Juillet 1975

(*) k sera pris égal à 100 ; s'il est besoin, on transformera la relation initiale $k_1 X + k_2Y + k_3Z = k$ en posant

$$x = \frac{100 \text{ X } k_1}{k}, \quad y = \frac{100 \text{ Y } k_2}{k}, \quad z = \frac{100 \text{ Z } k_3}{k}.$$

En pratique, x, y et z sont généralement positifs et inférieurs à 100 (cas de pourcentages).

La somme U des carrés des distances des points au premier axe principal, donc à $\Pi_2$, doit être minimale. Elle se note :

$$U = \sum_{i=1}^{n} (ax + by + cz - d)^2 .$$

Compte tenu de la relation (1), on trouve l'expression de $U_{min}$ en utilisant la méthode des multiplicateurs de *Lagrange :*

$$U_{min} = \lambda_{min} = \frac{(V_{11} + V_{22} + V_{33}) - \sqrt{(V_{11} + V_{22} + V_{33})^2 + 4\,(V_{12}^2 + V_{23}^2 + V_{13}^2 - V_{11} V_{22} - V_{11} V_{33} - V_{22} V_{33})}}{2}$$

$V_{ij}$ est la covariance de la variable i avec la variable j). $U_{min}$ est la plus petite valeur propre de la matrice des covariances. On sait que l'importance de l'axe se mesure par le rapport de la plus grande valeur propre à la trace de cette matrice, à savoir

$$1 - \frac{\lambda_{min}}{V_{11} + V_{22} + V_{33}} .$$

Posons $\alpha = V_{11} - \lambda_{min}$, $\beta = V_{12}$ et $\gamma = V_{13}$ ; les cœfficients a, b, c sont solution du système

$$\begin{cases} a + b + c = 0 \\ a^2 + b^2 + c^2 = 1 \\ \alpha a + \beta b + \gamma c = 0. \end{cases}$$

On obtient

$$a = \frac{\gamma - \beta}{\mu} \qquad b = \frac{\alpha - \gamma}{\mu} \qquad c = \frac{\beta - \alpha}{\mu},$$

avec

$$\mu = \sqrt{2\,(\alpha^2 + \beta^2 + \gamma^2 - \alpha\gamma - \alpha\beta - \beta\gamma)}.$$

D'autre part,

$$d = a\overline{x} + b\overline{y} + c\overline{z}$$

puisque l'axe principal passe par le centre de gravité des points.

## II – POSITIONNEMENT DE L'AXE DANS UN DIAGRAMME TRIANGULAIRE.

L'expression analytique du premier axe factoriel est donnée par le système

$$\begin{cases} ax + by + cz = d \\ x + y + z = k, \end{cases}$$

qui s''ecrit encore

$$\begin{cases} x = \dfrac{b - c}{a - b}\, z + \dfrac{d - bk}{a - b} \\[2ex] y = \dfrac{a - c}{b - a}\, z + \dfrac{d - ak}{b - a} \end{cases} \qquad (2)$$

Ce système caractérise une droite dans un diagramme triangulaire. Considérons une droite quelconque ($\Delta$) dans un diagramme ABC (cf. fig. 1). $X_0$, $Y_0$ et $Z_0$ sont les intersections de $\Delta$ avec respectivement les axes $\mathcal{A}_x, \mathcal{A}_y, \mathcal{A}_z$. Appliquons le théorème de *Thalès* à un point M quelconque de cette droite, dont les coordonnées sont $\overline{AP}$, $\overline{BQ}$ et $\overline{CR}$. Il vient :

$$\frac{\overline{AP}}{\overline{AL}} = \frac{\overline{Z_0M}}{\overline{Z_0Y_0}} = \frac{\overline{Z_0R}}{\overline{Z_0C}} = \frac{\overline{CR}}{\overline{Z_0C}} + 1$$

et

$$\frac{\overline{BQ}}{\overline{BY_0}} = \frac{\overline{X_0M}}{\overline{X_0Y_0}} = \frac{\overline{KR}}{\overline{KC}} = \frac{\overline{CR}}{\overline{KC}} + 1.$$

Par conséquent,

$$\begin{cases} \overline{AP} = \dfrac{\overline{AL}}{\overline{Z_0C}}\,\overline{CR} + \overline{AL} \\[2ex] \overline{BQ} = \dfrac{\overline{BY_0}}{\overline{KC}}\,\overline{CR} + \overline{BY_0}. \end{cases}$$

On retrouve formellement le système (2).

La droite $\Delta$ est commodément construite à partir de $\Omega$ $(\overline{x}, \overline{y}, \overline{z})$, qui est le centre de gravité, et

$$Y_0\left(\frac{d - bk}{a - b}, \frac{d - ak}{b - a}, 0\right),$$

qui est le point d'intersection de $\Delta$ avec $\mathcal{A}_y$.

On obtient la coordonnée d'un point du nuage sur l'axe en le projetant orthogonalement sur $\Delta$ dans le diagramme. Celui-ci résulte en effet d'une projection des trois axes de coordonnées sur $\Pi_2$, et la position relative de deux droites de $\Pi_2$ n'est pas modifiée par cette transformation. La droite qui passe par $M_1$ $(x_1, y_1, z_1)$ et est orthogonale à $\Delta$ a pour équations

$$\begin{cases} x = \dfrac{a}{c} z + x_1 - \dfrac{a}{c} z_1 \\[2ex] y = \dfrac{b}{c} z + y_1 - \dfrac{b}{c} z_1. \end{cases}$$

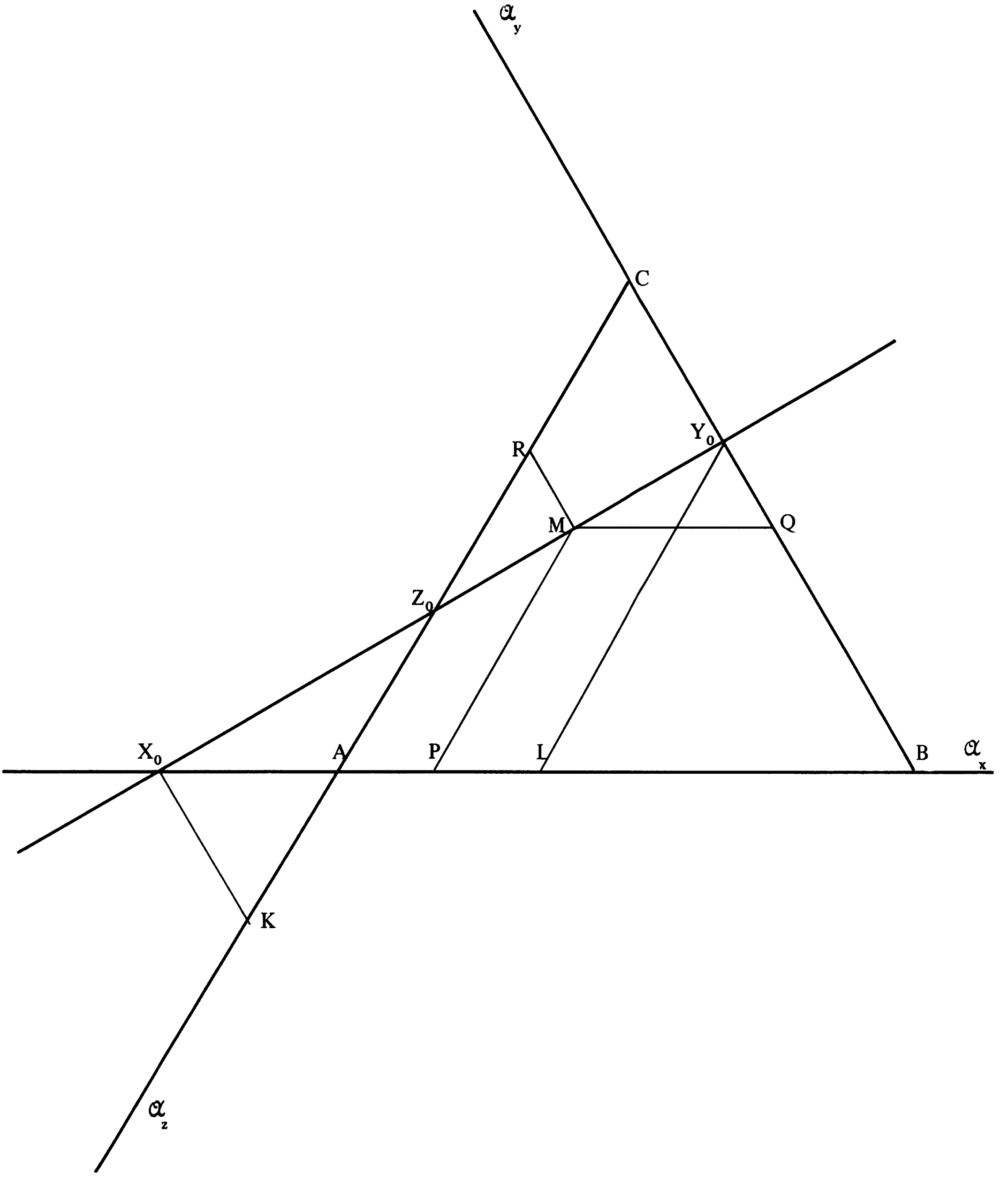


Figure 1 – Diagramme triangulaire (l'origine des coordonnées est en A, B, C)