Annales de l'I. H. P., section B

B. Garel

Tests de détection de valeurs aberrantes multidimensionnelles

Annales de l'I. H. P., section B, tome 14, nº 3 (1978), p. 303-314
<http://www.numdam.org/item?id=AIHPB_1978__14_3_303_0>

Ann. Inst. Henri Poincaré, Section B :
Vol. XIV, n° 3, 1978, p. 303-314. *Calcul des Probabilités et Statistique.*

Tests de détection de valeurs aberrantes multidimensionnelles

par

B. GAREL

Faculté des Sciences et Techniques de Chambéry, B. P. 143, 73011 Chambéry.
Laboratoire de Mathématiques Appliquées.
Université Scientifique et Médicale de Grenoble

RÉSUMÉ. — Nous proposons des tests de détection de valeurs aberrantes multidimensionnelles et démontrons l'optimalité de ces tests dans le cas d'un échantillon gaussien de matrice des variances-covariances Λ, supposée connue, et de glissements de la moyenne colinéaires.

SUMMARY. — Tests are given for outliers in a multinormal sample when the variance-covariance matrice Λ is known, and a property of optimality of these tests is shown when the observations have slipped in mean in parallel directions.

INTRODUCTION

Soit $x_1 \ldots x_n$ un n-échantillon gaussien multidimensionnel d'une loi normale $\mathcal{N}_p(m, \Lambda)$ à valeurs dans $\mathbb{R}^p$ de moyenne m inconnue et de matrice des variances-covariances Λ supposée connue. Notre étude se situe dans le cadre du modèle $\mathscr{A}$. C'est-à-dire que nous supposons qu'une valeur aberrante provient d'un glissement de la moyenne, soit d'une loi $\mathcal{N}_p(m + \Delta m, \Lambda)$, $\Delta m \in \mathbb{R}^p$, $\Delta m \neq 0$.

L'autre modèle, appelé modèle $\mathscr{B}$, consiste à supposer qu'une valeur aberrante provient d'une dilatation de la matrice des variances-covariances soit d'une loi $\mathcal{N}_p(m, \lambda\Lambda)$ avec λ réel > 1.

Si $x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ désigne un vecteur de $\mathbb{R}^p$ on notera x' le transposé de x.

1. FORMULATION DU PROBLÈME ET THÉORÈME PRINCIPAL

Soient $x_1 \dots x_n$ n variables aléatoires gaussiennes à valeurs dans $\mathbb{R}^p$ de moyenne $m_1, \dots, m_n$ respectivement et de même matrice des variances-covariances Λ connue. Nous voulons tester l'hypothèse

$$H_0 = H(m_1 = \dots = m_n = m)$$

contre les hypothèses alternatives que k ($k \leq n$) moyennes des observations ont glissé suivant des vecteurs de $\mathbb{R}^p$ colinéaires à un même vecteur $\Delta m \in \mathbb{R}^p$, k étant connu.

Nous noterons $\boldsymbol{\sigma}$ le groupe des permutations de l'ensemble $\{1, \dots, n\}$. Soit $\sigma \in \boldsymbol{\sigma}$, l'hypothèse H_σ est que

$$(1.0) \qquad \left| \begin{aligned} E(x_1) &= \alpha_{\sigma(1)}\Delta m + m \\ E(x_2) &= \alpha_{\sigma(2)}\Delta m + m \\ &\vdots \\ E(x_n) &= \alpha_{\sigma(n)}\Delta m + m \end{aligned} \right.$$

où $\alpha_1 \dots \alpha_n$ sont des scalaires réels connus. Nous avons donc $n!$ hypothèses alternatives. Les α_μ pour $\mu = 1, \dots, n$ ne sont pas nécessairement distincts et certains peuvent être nuls, l'un au moins étant différent de zéro. On donnera, par exemple, un résultat dans la situation $k = 2$, $\alpha_1 = \alpha_2 = 1$, $\alpha_\mu = 0$ pour $\mu = 3, \dots, n$.

Tout en particularisant H_0, nous formulons notre problème comme un problème de décision multiple dont nous présentons maintenant les principaux éléments, dans un cadre un peu plus général que nécessaire.

1° Soit $\mathscr{E}$ l'ensemble des états de la nature. Nous identifions $\mathscr{E}$ à l'espace produit $\mathscr{H} \times \mathbb{R}^p$ où $\mathscr{H}$ est l'ensemble discret des $r + 1$ hypothèses : $\mathscr{H} = \{H_0, H_1, \dots, H_r\}$ et où p est la dimension de l'espace d'un paramètre réel. $\mathbb{R}^p$ est muni de sa tribu des boréliens, $\mathscr{H}$ de son ensemble des parties $\mathscr{P}(\mathscr{H})$, et $\mathscr{E}$ de la tribu produit.

2° Nous disposons sur $\mathscr{E}$ d'une probabilité *a priori* que nous notons

$b \otimes \eta$ où b est une probabilité discrète *a priori* sur $\mathscr{H}$ et η une probabilité *a priori* absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}^p$: nous noterons $b[\mathrm{H}_\nu] = b_\nu$ la probabilité *a priori* de l'hypothèse H_ν, $\nu = 0, 1, \ldots, r$. On a donc $\sum_{\nu=0}^{r} b_\nu = 1$.

3° Nous noterons $\mathscr{D}$ l'ensemble discret des décisions possibles et de même cardinal que $\mathscr{H}$. $\mathscr{D} = \{ \mathrm{D}_0, \mathrm{D}_1, \ldots, \mathrm{D}_r \}$ où D_ν représente la décision d'accepter l'hypothèse H_ν, pour $\nu = 0, 1, \ldots, r$.

4° Nous noterons $f_\nu(x_1, \ldots, x_n)$ la densité conjointe des x_μ pour $\mu = 1, \ldots, n$ sous l'hypothèse H_ν, pour $\nu = 0, 1, \ldots, r$. De plus les f_ν pour $\nu = 0, 1, \ldots, r$ dépendent d'un paramètre vectoriel Δm à valeurs dans $\mathbb{R}^p$ sur lequel on dispose justement de la probabilité *a priori* η.

5° Nous noterons $\mathbb{D}$ l'ensemble des fonctions (ou règles) de décision aléatoires

$$\mathbb{D} = \{ d/d : (\mathbb{R}^p)^n \to \mathscr{M} \}$$

où $\mathscr{M}$ est un ensemble de probabilités sur $\mathscr{D}$. On peut identifier d (voir [5]) à une application $d : (\mathbb{R}^p)^n \to [0, 1]^{r+1}$

$$d = (d_0, d_1, \ldots, d_r) \qquad \text{avec } 0 \leq d_\nu(x_1, \ldots, x_n) \leq 1$$

$\sum_{\nu=0}^{r} d_\nu(x_1, \ldots, x_n) = 1$; $d_\nu(x_1, \ldots, x_n)$ représente la probabilité de prendre la décision D_ν sachant $(x_1, \ldots, x_n)$. Nous noterons : $\underline{x} = (x_1, \ldots, x_n)$ et $d\underline{x} = dx_1 \ldots dx_n$. De plus pour $\nu = 0, 1, \ldots, r$ nous supposerons que $d_\nu(\underline{x})$ est mesurable par rapport à la mesure de Lebesgue sur $\mathbb{R}^{n \cdot p}$.

6° Nous appellerons L la fonction de coût. L est une application de $\mathscr{H} \times \mathscr{D} \to \mathbb{R}^+$. $\mathrm{L}(\mathrm{D}_\nu/\mathrm{H}_\xi)$ est le coût de la décision D_ν lorsque l'état H_ξ est réalisé. Dans ce qui suit, nous prendrons la fonction L définie par

$$\mathrm{L}(\mathrm{D}_\nu/\mathrm{H}_\xi) = \begin{cases} 0 & \text{si} \quad \nu = \xi \\ 1 & \text{si} \quad \nu \neq \xi \end{cases}$$

Il est alors possible de définir le risque de Bayes et les règles de Bayes relatifs à la probabilité *a priori* $b \otimes \eta$, et de caractériser ces dernières. Nous renvoyons pour une étude détaillée à [*1*]. Notons que dans le cas de glissements colinéaires décrit ci-dessus $r = n!$. On notera D_0 (resp. D_σ) la décision de considérer H_0 (resp. H_σ) comme réalisé et $d_0(\underline{x})$ (resp. $d_\sigma(\underline{x})$) la probabilité de prendre la décision D_0 (resp. D_σ) sachant $\underline{x}$.

Nous définissons sur $\mathscr{H}$ la probabilité *a priori* b telle que

$$b[H_0] = b_0$$

$$b[H_\sigma] = \frac{1 - b_0}{n!}$$

Soit $c > 0$ une constante fixée telle que $\| \Delta m \|^2_{\Lambda^{-1}} = c$, où $\| \Delta m \|^2_{\Lambda^{-1}} = \Delta m' \Lambda^{-1} \Delta m$.

Nous définissons la probabilité *a priori* $n_c(\Delta m)$ comme étant la mesure concentrée sur l'ellipsoïde $\| \Delta m \|^2_{\Lambda^{-1}} = c$ telle qu'en le transformant par une application linéaire en la sphère unité S_1 de $\mathbb{R}^p$ la mesure image obtenue soit la mesure uniforme $dS(\Delta m)$ sur la sphère unité.

On choisit de se limiter aux règles de décision possédant les propriétés suivantes :

1.1. La probabilité de choisir D_0 sous l'hypothèse H_0 est

$$P(D_0/d, H_0) = 1 - a,$$

niveau de confiance du test.

Donc a est fixé, et ne varie pas avec d.

1.2. Les règles de décision sont invariantes lorsqu'on ajoute un vecteur constant $y \in \mathbb{R}^p$ à chaque observation :

$$\forall y \in \mathbb{R}^p, \quad \forall \sigma \in \boldsymbol{\sigma} \qquad d_\sigma(x_1 + y, \ldots, x_n + y) = d_\sigma(x_1 \ldots x_n)$$

et

$$d_0(x_1 + y, \ldots, x_n + y) = d_0(x_1 \ldots x_n)$$

Les règles de décisions auxquelles on se limite sont donc invariantes sur un sous-groupe, noté G_1, du groupe des translations de $\mathbb{R}^{n \cdot p}$.

1.3. Enfin, les règles de décision doivent vérifier : les probabilités de décision $P(D_\sigma/d, (H_\sigma, \Delta m))$ de choisir D_σ sous l'hypothèse H_σ sont indépendantes de $\sigma \in \boldsymbol{\sigma}$, et ne dépendent que de $\| \Delta m \|_{\Lambda^{-1}}$.

1.4. THÉORÈME. — Dans le cadre du modèle décrit ci-dessus, la règle de décision possédant les propriétés 1.1, 1.2 et 1.3 qui maximise les probabilités de décision $P(D_\sigma/d, (H_\sigma, \Delta m))$ uniformément en $\Delta m \in \mathbb{R}^p$ et $\sigma \in \boldsymbol{\sigma}$, nous est donnée par :

$$\left| \begin{array}{ll} \text{si} \quad R^2_{(\sigma)} < \lambda_a & \text{prendre la décision } D_0 \\ \text{si} \quad R^2_{(\sigma)} > \lambda_a & \text{prendre la décision } D_{(\sigma)} \end{array} \right.$$

où (σ) est la permutation pour laquelle R^2_σ est maximal, où

$$R^2_\sigma = \left[\sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_\mu - \bar{x}) \right]' \Lambda^{-1} \left[\sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_\mu - \bar{x}) \right]$$

où λ_a est une constante qui dépend du niveau de signification, donc de a, avec $P_{H_0}(R^2_{(\sigma)} < \lambda_a) = 1 - a$.

1.5. *Remarque.* — La probabilité $P(D_\sigma/d, (H_\sigma, \Delta m))$ étant indépendante de $\sigma \in \boldsymbol{\sigma}$ on l'appelle probabilité de détection de k valeurs aberrantes et l'on dit que la règle du théorème 1.4 maximise uniformément en Δm la probabilité de détection de k valeurs aberrantes éventuelles.

Preuve du théorème : les probabilités de décision d'une règle de décision vérifiant l'invariance sous le groupe G_1 sont indépendantes de m, la moyenne inconnue. Nous supposerons donc $m = 0$.

Alors (voir [*1*]) parmi les règles de décision possédant les propriétés 1.1, 1.2 et 1.3, celles qui maximisent uniformément en $\sigma \in \boldsymbol{\sigma}$ et $\Delta m \in \mathbb{R}^p$ les probabilités de décision $P(D_\sigma/d_\sigma, (H_\sigma, \Delta m))$ sont les règles de Bayes relatives à la probabilité *a priori* $b \otimes \eta_c$. D'autre part, les règles de Bayes relatives à la probabilité *a priori* $b \otimes \eta_c$ possédant les propriétés 1.1 et 1.2, c'est-à-dire de niveau de signification $1 - a$ et invariantes sous le groupe de transformations G_1 sont données par

$$
(1.6)\quad \left|\begin{array}{lll}
d_0(\underline{x}) = 1 & \text{si} & A h_0 > \quad h_\sigma(\underline{x}, \Delta m) d\eta_c(\Delta m) \qquad \forall \sigma \in \boldsymbol{\sigma} \\[2ex]
d_\sigma(\underline{x}) = 1 & \text{si} & \displaystyle\int h_\sigma(\underline{x}, \Delta m) d\eta_c(\Delta m) > A h_0 \\[2ex]
 & \text{et} & \displaystyle\int h_\sigma(\underline{x}, \Delta m) d\eta_c(\Delta m) > \int h_{\sigma'}(\underline{x}, \Delta m) d\eta_c(\Delta m) \\
 & & \qquad\qquad \forall \sigma' \in \boldsymbol{\sigma} \qquad \text{avec } \sigma' \neq \sigma
\end{array}\right.
$$

sauf peut-être sur un sous-ensemble E de $\mathbb{R}^{n \cdot p}$ de mesure nulle. h_0 (resp. h_σ) est la densité de l'invariant maximal sous le groupe G_1 exprimé en fonction de $\underline{x}$, sous l'hypothèse H_0 (resp. H_σ). A est une constante de $\mathbb{R}^{+*}$ qui dépend de a.

Par un sous-ensemble $E \subset \mathbb{R}^{np}$ de mesure nulle nous voulons dire un ensemble E tel que

$$\int\int_E f_\sigma(\underline{x}, \Delta m) d\underline{x}\, d(b \otimes \eta_c) = 0 .$$

L'invariant maximal sous le groupe G_1 nous est donné par une généralisation en dimension p de l'exemple 1, page 216 de [*3*], résultat que l'on peut trouver dans [*1*], page 53, soit : $x_1 - x_n, \ldots, x_{n-1} - x_n$.

Nous avons donc besoin de calculer la densité de l'invariant maximal sous le groupe G_1 sous les hypothèses H_0 et H_σ, $\sigma \in \boldsymbol{\sigma}$. Ces résultats sont des corollaires d'un lemme démontré dans [*2*] que nous rappelons ici.

1.7. LEMME. — Soient $x_1 \dots x_n$, n variables aléatoires gaussiennes à valeurs dans $\mathbb{R}^p$ de moyennes respectives $m_1 \dots m_n$ et de même matrice des variances-covariances Λ. La densité conjointe des $(n-1)$ variables $x_1 - x_n, \dots, x_{n-1} - x_n$ que nous notons $h(x_1, \dots, x_n)$ nous est donnée par

$$h(x_1 \dots x_n) = C_0 \exp\left\{-\frac{1}{2n} Q\right\} \tag{1.8}$$

où

$$Q = \sum_{ij} \lambda^{ij} \left\{ n \sum_{\mu=1}^{n} (x_{\mu i} - m_{\mu i})(x_{\mu j} - m_{\mu j}) - \left[\sum_{\mu=1}^{n} (x_{\mu i} - m_{\mu i}) \sum_{\mu=1}^{n} (x_{\mu j} - m_{\mu j}) \right] \right\}$$

et (λ^{ij}) est la matrice inverse de la matrice Λ, et où

$$C_0 = \frac{1}{n^{p/2}[(2\pi)^p \,|\Lambda|]^{\frac{n-1}{2}}},$$

en notant $|\Lambda|$ le déterminant de la matrice Λ.

1.9. COROLLAIRE. — Sous l'hypothèse H_0 la densité de l'invariant maximal $x_1 - x_n, \dots, x_{n-1} - x_n$ nous est donnée par

$$h_0(x_1 \dots x_n) = C_0 \exp\left\{ -\frac{1}{2} \sum_{\mu=1}^{n} (\Delta x_\mu)' \Lambda^{-1} (\Delta x_\mu) \right\}$$

où

$$\Delta x_\mu = x_\mu - \bar{x}$$

(*Preuve.* — En traduisant l'hypothèse $m_1 = \dots = m_n = 0$ on trouve le résultat à partir de la formule (1.8) du lemme 1.7).

1.10. COROLLAIRE. — Sous l'hypothèse H_σ, $\sigma \in \boldsymbol{\sigma}$ définie en (1.0), la densité de l'invariant maximal $x_1 - x_n, \dots, x_{n-1} - x_n$ s'écrit

$$h_\sigma(x_1, \dots, x_n, \Delta m) = h_0 \exp\left\{ c_1 \,\|\Delta m\|^2_{\Lambda^{-1}} + \left\langle \Delta m, \sum_{\mu=1}^{n} \alpha_{\sigma(\mu)} \Delta x_\mu \right\rangle_{\Lambda^{-1}} \right\}$$

où $\langle \cdot, \cdot \rangle_{\Lambda^{-1}}$ est le produit scalaire associé à Λ^{-1} ; et

$$c_1 = \frac{1}{2n}\left[\left(\sum_{\mu=1}^{n} \alpha_{\sigma(\mu)} \right)^2 - n \left(\sum_{\mu=1}^{n} \alpha^2_{\sigma(\mu)} \right) \right]$$

Preuve. — Par hypothèse pour $\mu = 1, \ldots, n$

$$m_\mu = \alpha_{\sigma(\mu)}\Delta m$$

Q s'écrit :

$$\begin{aligned}
Q = \sum_{ij} \lambda^{ij} \Bigg\{ n \sum_{\mu=1} (x_{\mu i} - \alpha_{\sigma(\mu)}\Delta m_i)(x_{\mu j} - \alpha_{\sigma(\mu)}\Delta m_j) \\
- \left[\sum_{\mu=1}^{n} (x_{\mu i} - \alpha_{\sigma(\mu)}\Delta m_i) \sum_{\mu=1}^{n} (x'_{\mu j} - \alpha_{\sigma(\mu)}\Delta m_j) \right] \Bigg\} \\
= \sum_{ij} \lambda^{ij} \Bigg\{ n \sum_{\mu=1}^{n} x_{\mu i}x_{\mu j} - n\Delta m_j \sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}x_{\mu i} \\
- n\Delta m_i \sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}x_{\mu j} + n\Delta m_i\Delta m_j \sum_{\mu=1}^{n} \alpha^2_{\sigma(\mu)} \\
- n^2\bar{x}_i\bar{x}_j + n\bar{x}_j\Delta m_i \sum_{\mu=1}^{n} \alpha_{\sigma(\mu)} + n\bar{x}_i\Delta m_j \sum_{\mu=1}^{n} \alpha_{\sigma(\mu)} \\
- \left(\sum_{\mu=1}^{n} \alpha_{\sigma(\mu)} \right)^2 \Delta m_i\Delta m_j \Bigg\}
\end{aligned}$$

$$\begin{aligned}
h_\sigma(x_1, \ldots, x_n, \Delta m) = h_0 \exp \Bigg\{ -\frac{1}{2n} \left[\sum_{ij} \lambda^{ij}\Delta m_i\Delta m_j \left(n \sum_{\mu=1}^{n} \alpha^2_{\sigma(\mu)} - \left(\sum_{\mu=1}^{n} \alpha_{\sigma(\mu)} \right)^2 \right) \right] \\
+ \sum_{ij} \lambda^{ij} \left[\Delta m_j \sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_{\mu i} - \bar{x}_i) \right] \Bigg\}
\end{aligned}$$

en tenant compte de nos notations

$$Q = h_0 \exp \left\{ C_1\Delta m'\Lambda^{-1}\Delta m + \Delta m'\Lambda^{-1} \left[\sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_\mu - \bar{x}) \right] \right\}$$

on en conclut le corollaire 1.10.)

Pour terminer la démonstration du théorème 1.4, posons

$$Z_\sigma = \sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_\mu - \bar{x}).$$

Nous nous proposons de calculer les intégrales de la formule (1.6) soit :

$$\mathrm{I}_\sigma = \int_{\|\Delta m\|^2_{\Lambda^{-1}}} h_\sigma(x_1, \ldots, x_n, \Delta m) d\eta_c(\Delta m)$$

où h_σ nous est donnée par le corollaire 1.10.

$$\mathrm{I}_\sigma = h_0 \int_{\|\Delta m\|^2_{\Lambda^{-1}}} \exp \{ c_1 \| \Delta m \|^2_{\Lambda^{-1}} + \langle \Delta m, \mathrm{Z}_\sigma \rangle_{\Lambda^{-1}} \} d\eta_c(\Delta m)$$

Soit D la matrice associée à l'application linéaire qui transforme l'ellipsoïde $\| \Delta m \|^2_{\Lambda^{-1}} = c$ en la sphère unité $\| \Delta m \|^2 = 1$ où $\| \, . \, \|$ désigne la norme euclidienne de $\mathbb{R}^p$. Si $\mathscr{I}_p$ désigne la matrice identité d'ordre p nous avons

$$\mathscr{I}_p = \mathrm{D}'\Lambda^{-1}\mathrm{D} \Leftrightarrow \Lambda^{-1} = (\mathrm{D}')^{-1}\mathrm{D}^{-1}$$

Dans l'intégrale opérons le changement de variable $\Delta m \to \mathrm{D}\Delta m$ et appliquons le résultat sur la mesure image. L'intégrale devient

$$\mathrm{I}_\sigma = h_0 e^{c_1} \int_{\|\Delta m\|^2 = 1} \exp \{ \langle \Delta m, \mathrm{D}^{-1}\mathrm{Z}_\sigma \rangle \} d\mathrm{S}(\Delta m)$$

Opérons alors un changement de base orthogonale de telle sorte que dans le nouveau système de coordonnées le vecteur $\mathrm{D}^{-1}\mathrm{Z}_\sigma$ aient ses $(p - 1)$ premières composantes nulles et la p^{o} strictement positive. La distribution $d\mathrm{S}(\Delta m)$ étant uniforme nous avons

$$\begin{aligned}
\mathrm{I}_\sigma &= h_0 e^{c_1} \int_{\|\Delta m\|^2 = 1} \exp (\Delta m_p \| \mathrm{D}^{-1}\mathrm{Z}_\sigma \|) d\mathrm{S}(\Delta m) \\
&= h_0 e^{c_1} \int_{\substack{\|\Delta m\|^2 = 1 \\ \text{et } \Delta m_p > 0}} \exp (\Delta m_p \| \mathrm{D}^{-1}\mathrm{Z}_\sigma \|) d\mathrm{S}(\Delta m) \\
&\quad + h_0 e^{c_1} \int_{\substack{\|\Delta m\|^2 = 1 \\ \text{et } \Delta m_p > 0}} \exp (- \Delta m_p \| \mathrm{D}^{-1}\mathrm{Z}_\sigma \|) d\mathrm{S}(\Delta m) \\
&= 2 h_0 e^{c_1} \int_{\substack{\|\Delta m\|^2 = 1 \\ \text{et } \Delta m_p > 0}} \mathrm{ch} (\Delta m_p \| \mathrm{D}^{-1}\mathrm{Z}_\sigma \|) d\mathrm{S}(\Delta m)
\end{aligned}$$

pour $\Delta m_p > 0$, $\mathrm{ch} (\Delta m_p \| \mathrm{D}^{-1}\mathrm{Z}_\sigma \|)$ est une fonction strictement croissante de $\| \mathrm{D}^{-1}\mathrm{Z}_\sigma \|$. Il en est de même de I_σ et l'on a

$$\begin{aligned}
\mathrm{A}h_0 > \mathrm{I}_\sigma \Leftrightarrow \mathrm{A} > 2e^{c_1} &\int_{\substack{\|\Delta m\|^2 = 1 \\ \text{et } \Delta m_p > 0}} \mathrm{ch} (\Delta m_p \| \mathrm{D}^{-1}\mathrm{Z}_\sigma \|) d\mathrm{S}(\Delta m) \\
&\Leftrightarrow \| \mathrm{D}^{-1}\mathrm{Z}_\sigma \|^2 < \lambda_a, \qquad \text{pour tout } \sigma \in \boldsymbol{\sigma}
\end{aligned}$$

Choisir $\lambda_a > 0$ de telle sorte que la propriété 1.1 soit vérifiée revient à déterminer $A \in \mathbb{R}^{*+}$.

D'autre part pour σ et $\sigma' \in \boldsymbol{\sigma}$

$$I_\sigma > I_{\sigma'} \Leftrightarrow \int_{\substack{\|\Delta m\|^2=1 \\ \text{et } \Delta m_p > 0}} \operatorname{ch}(\Delta m_p \| D^{-1} Z_\sigma \|) dS(\Delta m)$$

$$> \int_{\substack{\|\Delta m\|^2=1 \\ \text{et } \Delta m_p > 0}} \operatorname{ch}(\Delta m_p \| D^{-1} Z_\sigma \|) dS(\Delta m) \Leftrightarrow \| D^{-1} Z_\sigma \| > \| D^{-1} Z_{\sigma'} \|$$

Remarquons que

$$\begin{aligned} \| D^{-1} Z_\sigma \|^2 &= Z'_\sigma (D^{-1})' D^{-1} Z_\sigma = Z'(D')^{-1} D^{-1} Z_\sigma \\ &= Z'_\sigma \Lambda^{-1} Z_\sigma \end{aligned}$$

Relativement à la distribution *a priori* $\eta_c(\Delta m)$ les règles de Bayes possédant les propriétés 1.1 et 1.2 sont caractérisées par

$$(1.11) \quad \left| \begin{array}{l} d_0(\underline{x}) = 1 \quad \text{si} \\ R^2_{(\sigma)} = \max\limits_{\sigma \in \boldsymbol{\sigma}} \left[\sum\limits_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_\mu - \bar{x}) \right]' \Lambda^{-1} \left[\sum\limits_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_\mu - \bar{x}) \right] < \lambda_a \\ d_{(\sigma)}(x) = 1 \quad \text{si} \quad R^2_{(\sigma)} > \lambda_a \end{array} \right.$$

Cette règle de décision est manifestement symétrique en $x_1 \ldots x_n$. Les densités des lois conjointes des variables $x_1 \ldots x_n$ sous les diverses hypothèses sont également symétriques en $x_1 \ldots x_n$. Enfin pour tout $\sigma \in \boldsymbol{\sigma}$ on vérifie aisément que les probabilités de décision $P(D_\sigma/d, (H_\sigma, \Delta m))$ de la règle d ci-dessus ne dépendent que de $\| \Delta m \|_{\Lambda^{-1}}$. La règle de décision (1.11) possède donc la propriété 1.3 ce qui termine la démonstration du théorème 1.4.

2. EXEMPLES D'APPLICATIONS

Nous donnons ici trois corollaires qui nous permettent de retrouver le résultat connu et, démontré dans [*2*] et de généraliser certains résultats en dimension 1, présentés dans [*4*].

2.1. COROLLAIRE. — Sous l'hypothèse qu'il n'y a au plus qu'une valeur aberrante, dans le cadre du modèle décrit au paragraphe 1, la règle de décision possédant les propriétés 1.1, 1.2 et 1.3 qui maximise la probabilité de détection de cette valeur uniformément en $\Delta m \in \mathbb{R}^p$ nous est donnée par :

$$\left| \begin{array}{ll} \text{prendre la décision } D_0 & \text{si } \chi^2_{(n)} = \max\limits_{\mu=1\ldots n} (x_\mu - \bar{x})'\Lambda^{-1}(x_\mu - \bar{x}) < \lambda_a \\ \text{prendre la décision } D_{(n)} & \text{si } \chi^2_{(n)} > \lambda_a \end{array} \right.$$

(*Preuve.* — Dans ce cas, un seul des α_μ est non nul, égal à 1 et il y a n hypothèses alternatives).

Ce résultat est déjà connu et dû à [*2*].

2.2. Corollaire. — Supposons que $1 < k < n$ des α_μ soient non nuls, tous égaux à 1. Alors la règle de décision satisfaisant aux hypothèses du théorème 1.4 et possédant les mêmes propriétés nous est donnée par :

$$\left| \begin{array}{ll} \text{prendre la décision } D_0 & \text{si } R^2_{(\sigma)} < \lambda_a \\ \text{prendre la décision } D_{(\sigma)} & \text{si } R^2_{(\sigma)} > \lambda_a \end{array} \right.$$

où

$$R^2_{(\sigma)} = \max_{\sigma \in \boldsymbol{\sigma}} \left[\sum_{\mu=1}^{k} (x_{\sigma(\mu)} - \bar{x}) \right]' \Lambda^{-1} \left[\sum_{\mu=1}^{k} (x_{\sigma(\mu)} - \bar{x}) \right]$$

et λ_a est une constante dépendant de a.

(*Preuve.* — Sans perdre de généralité, nous pourrons supposer

$$\begin{aligned} \alpha_1 = \ldots = \alpha_k = 1 \\ \alpha_{k+1} = \ldots = \alpha_n = 0 \end{aligned}$$

alors

$$\sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_\mu - \bar{x}) = \sum_{\mu=1}^{n} \alpha_\mu (x_{\sigma^{-1}(\mu)} - \bar{x})$$

Alors

$$\max_{\sigma \in \boldsymbol{\sigma}} \left[\sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_\mu - \bar{x}) \right]' \Lambda^{-1} \left[\sum_{\mu=1}^{n} \alpha_{\sigma(\mu)}(x_\mu - \bar{x}) \right]$$

$$= \max_{\sigma^{-1} \in \boldsymbol{\sigma}} \left[\sum_{\mu=1}^{n} \alpha_\mu (x_{\sigma^{-1}(\mu)} - \bar{x}) \right]' \Lambda^{-1} \left[\sum_{u=1}^{n} \alpha_\mu (x_{\sigma^{-1}(\mu)} - \bar{x}) \right]$$

$$= \max_{\sigma \in \boldsymbol{\sigma}} \left[\sum_{\mu=1}^{n} \alpha_\mu (x_{\sigma(\mu)} - \bar{x}) \right] \Lambda^{-1} \left[\sum_{\mu=1}^{n} \alpha_\mu (x_{\sigma(\mu)} - \bar{x}) \right]$$

en remplaçant les α_μ par leur valeur on obtient le résultat.)

En particulier lorsque $k = 2$, $\alpha_1 = \alpha_2 = 1$, $\alpha_\mu = 0$ pour $\mu = 3, \ldots, n$ on trouve le résultat simple suivant : le test de détection optimal est associé à la statistique

$$\max_{1 \leq i < j \leq n} (x_i + x_j - 2\bar{x})'\Lambda^{-1}(x_i + x_j - 2\bar{x})$$

Supposons enfin que les n observations soient partagées en r populations de k éléments : $n = rk$.

Nous nous proposons de savoir si l'une des populations a glissé en moyenne.

2.3. Corollaire. — La règle de décision correspondante, vérifiant les hypothèses du théorème 1.4 et possédant les propriétés qui y sont énoncées, nous est donnée par prendre la décision

$$\begin{array}{lll} D_0 & \text{si} & R^2_{(l)} < \lambda'_a \\ D_{(l)} & \text{si} & R^2_{(l)} > \lambda'_a \end{array}$$

où (l) désigne l'indice pour lequel

$$R^2_l = (\bar{x}(l) - \bar{x})'\Lambda^{-1}(\bar{x}(l) - \bar{x}) \qquad \text{est maximal}; \ 1 \leq l \leq r$$

$$\lambda'_a = \lambda_a/k^2$$

où

$$\bar{x}(l) = \frac{1}{k}\sum_{j=1}^{k} X_{\mu(j)}$$

(*Preuve.* — Par hypothèse k des coefficients sont non nuls, égaux à 1, les autres étant nuls.

Les seules permutations qui présentent un intérêt sont celles pour lesquelles les coefficients $\alpha_\mu \neq 0$ opèrent sur les éléments d'un même groupe. Nous avons donc r hypothèses alternatives, et l'on a :

$$R^2_{\sigma(l)} = \left[\sum_{j=1}^{k} (x_{\mu(j)} - \bar{x}\right]'\Lambda^{-1}\left[\sum_{j=1}^{k} (x_{\mu(j)} - \bar{x}\right]$$

où $\mu(j)$ désigne un des k indices μ pour lesquels $\alpha_\mu = 1$, et $\sigma(l)$ désigne la l° des r permutations présentant un intérêt.

$$\begin{aligned} R^2_{\sigma(l)} &= \left[\sum_{j=1}^{k} x_{\mu(j)} - k\bar{x}\right]'\Lambda^{-1}\left[\sum_{j=1}^{k} x_{\mu(j)} - k\bar{x}\right] \\ &= k^2(\bar{x}(l) - \bar{x})'\Lambda^{-1}(\bar{x}(l) - \bar{x}) \end{aligned}$$

En posant $R^2_l = R^2_{\sigma(l)}$, on obtient le résultat.)

Bien que l'on ne connaisse pas Δm, on a supposé les α_μ connus. Nous pouvons donc dire qu'en l'absence de toute information sur la nature des glissements il n'y a pas de règle optimale de détection de $k > 1$ valeurs aberrantes.

RÉFÉRENCES

[1] GAREL, Détection des valeurs aberrantes dans un échantillon gaussien multidimensionnel. *Thèse de 3e cycle présentée à Grenoble*, 1976, p. 39-62.

[2] KUDO, The extreme value in a multivariate normal sample. *Mem. fac. Sci. Kyushu Univ.* (A), t. **11**, 1957, p. 143-156.

[3] LEHMAN, *Testing statistical hypotheses.* Wiley and Sons, New York, 1959.

[4] MacMILLAN, *Tests for one or two outliers.* Ph. D. Thesis, North Carolina state Univ., 1968.

[5] WALD, *Statistical decision functions.* Chelsea. Publ. Company, New York, 1971.

(Manuscrit reçu le 1er février 1978)