2, 4. Ecarts au modèle

pas symétriques, à moins que $\mathsf{U}_q^\star$ et $\mathsf{U}_{r-q}^\star$ ne soient mutuellement orthogonaux; dans ce dernier cas,

$$\mathbf{red}\left[\mathsf{U}_q^\star\right] = \mathbf{red}\left[\mathsf{U}_q^\star \;\middle|\; \mathsf{U}_{r-q}^\star\right] = \mathbf{SC}\,\mathsf{U}_q^\star \,,$$
$$\mathbf{red}\left[\mathsf{U}_{r-q}^\star\right] = \mathbf{red}\left[\mathsf{U}_{r-q}^\star \;\middle|\; \mathsf{U}_q^\star\right] = \mathbf{SC}\,\mathsf{U}_{r-q}^\star \,.$$

2, 33. Ces considérations s'étendent aisément au cas où V_+ est décomposé en plus de deux sous-espaces, suivant le schéma

$$\begin{cases} \mathsf{V}_+ = \mathsf{U}_1^\star \oplus \mathsf{U}_2^\star \oplus \dots \oplus \mathsf{U}_t^\star \,, \\ \dim \mathsf{U}_i^\star = r_i \,, \quad r_1 + \dots + r_k = \rho_k ;\ \rho_t = r \,. \end{cases}$$

On doit alors considérer t modèles successifs (et l'ordre dans lequel ces modèles font intervenir les $\mathsf{U}_i^\star$ est essentiel); le $k^{\text{ème}}$ de ces modèles est caractérisé par

$$\left[\mathfrak{l}^\star \in \bigoplus_{k+1}^{t} \mathsf{U}_i^\star\right] \quad \text{implique} \quad \mathbf{E}\,\mathfrak{l}^\star \mathfrak{x} = 0 \qquad (k = 1, 2, \dots, t-1) \,,$$

le $t^{\text{ème}}$ étant le modèle initial. On note $\mathbf{SCE}_{n-\rho_k}$ la somme de carrés des erreurs attachée au $k^{\text{ème}}$ modèle, et on montre sans peine que

$$\mathbf{SC}\left(\bigoplus_{k+1}^{t} \mathsf{U}_i^\star\right) = \mathbf{SCE}_{n-\rho_k} - \mathbf{SCE}_{n-r} \,;$$

on pose alors

$$\mathbf{red}\left[\mathsf{U}_1^\star\right] = \mathbf{SCT} - \mathbf{SCE}_{n-\rho_1}$$
$$\mathbf{red}\left[\mathsf{U}_{k+1}^\star \;\middle|\; \mathsf{U}_1^\star, \dots, \mathsf{U}_k^\star\right] = \mathbf{SCE}_{n-\rho_k} - \mathbf{SCE}_{n-\rho_{k+1}} \,,$$

et on a

$$\mathbf{SCN} = \mathbf{red}\left[\mathsf{U}_1^\star\right] + \sum_{1}^{t-1} \mathbf{red}\left[\mathsf{U}_{k+1}^\star \;\middle|\; \mathsf{U}_1^\star, \dots, \mathsf{U}_k^\star\right] , \tag{12}$$

avec

$$\mathbf{red}\left[\mathsf{U}_t^\star \;\middle|\; \mathsf{U}_1^\star, \dots, \mathsf{U}_{t-1}^\star\right] = \mathbf{SC}\,\mathsf{U}_t^\star \,,$$

cette dernière relation n'étant pas généralement vraie pour les autres $\mathsf{U}_i^\star$ (exception évidente: le cas où les $\mathsf{U}_i^\star$ sont mutuellement orthogonaux).

2, 4. *Ecarts au modèle.*

Tout ce qui précède est valide si, réellement, $\mathbf{E}\,\mathfrak{x} = \mathfrak{A}\mathfrak{b}$. S'il n'en est pas nécessairement ainsi, ce qui arrive lorsque le modèle envisagé n'est qu'un cas particulier d'un modèle plus général auquel on désire accorder aussi quelque considération, on peut modifier un peu les énoncés des hypothèses à éprouver, en disant, par exemple: « *si* $\mathbf{E}\mathfrak{x} = \mathfrak{A}\mathfrak{b}$ et *si* $\mathfrak{l}_1^\star \mathfrak{b} = \ldots = \mathfrak{l}_s^\star \mathfrak{b} = 0$, alors ... ». Sous cette nouvelle hypothèse, $\mathbf{SC}\{\mathfrak{l}_1^\star \hat{\mathfrak{b}}, \ldots, \mathfrak{l}_s^\star \hat{\mathfrak{b}}\}$ est encore distribuée comme $\sigma^2 \chi_s^2$. Mais **SCE** n'est plus distribuée comme $\sigma^2 \chi_{n-r}$, car, si $\mathbf{E}\mathfrak{x} - \mathfrak{A}\mathfrak{b}$ n'est pas identiquement nulle, les vecteurs de V_0 n'ont plus une moyenne nécessairement nulle. On est alors obligé de prendre comme espace des erreurs un sous-espace V_* de V_0, à savoir: celui des vecteurs de $\mathsf{V}^\star$ dont la moyenne est identiquement nulle dans le modèle le plus général que l'on considère. On peut dire que ce sous-espace existe dès que les observations comportent au moins une paire d'observations ayant identiquement même moyenne (dans le modèle le plus général). Nous noterons **SCint** (« somme de carrés interne ») l'expression $\mathbf{SC}\,\mathsf{V}_*$, et **SCEM** [« somme de carrés des écarts au modèle » (sous-entendu: au modèle restreint)] l'expression **SCE** — **SCint** (en désignant par V'_* le complément orthogonal de V_* dans V_0, **SCEM** n'est autre que $\mathbf{SC}\,\mathsf{V}'_*$). Dans les considérations du § 2, 2, **SCint** peut remplacer **SCE**, $n - s$ remplaçant alors $n - r$. Les composantes de **SCint** sont évidemment orthogonales à celles de **SCN**.

Remarque. — Les composantes additives de **SCT** (ou, plus exactement, leurs valeurs observées) sont le plus souvent reprises en un tableau que l'on nomme « table d'analyse de la variance ». Cette désignation n'est guère heureuse, on devrait la réserver aux études de « composantes de variance » (cfr. [V]); elle paraît néanmoins avoir reçu la sanction de l'usage, et il semble assez vain de vouloir la récuser. Une telle table se présente ainsi:

Somme de carrés	Formules	Nombre de degrés de liberté
SCT	$\sum x_i^2$	n
SCN	$SC\{\mathfrak{A}^T x\}$	r
SCE	$SCT - SCN$	$n - r$
$\quad\{$ $SCint$	$SC\mathsf{V}_*$	u
$\quad\{$ $SCEM$	$SCE - SCint$	$n - r - u$.

D'ordinaire, *SCN* est décomposée conformément à la formule (12). Si on utilise la table de la distribution **F**, il est utile d'adjoindre à cette table une colonne « carrés moyens », où sont repris les quotients des *SC* par les nombres de leurs degrés de liberté.

NOTES

1) Dans un système complet de notations, ce n-uple serait désigné, par exemple, par $\mathfrak{x}_p$.

2) Dans un système complet de notations, ce n-uple serait désigné par $\mathfrak{x}^{\star}_{p\star}$, ou par $\mathfrak{x}^{\star}_p$ si la dualité des bases va de soi. Pour des raisons de convenance typographique, nous écrirons souvent $[a_1, \ldots, a_n]^{\star}$ au lieu de $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$

3) Ce second usage est permis parce que $\mathfrak{P}$ et $\mathfrak{P}^{\star}$ sont des bases orthonormales; s'il n'en était pas ainsi, il conviendrait d'indiquer la transposition et la dualité par des signes différents (^{T} et *).

4) On notera qu'alors $\mathfrak{A}^T$ ne représente pas l'application duale (dite aussi « transposée ») de $\mathfrak{A}$; celle-ci est représentée, ici, par la même matrice $\mathfrak{A}$; mais, dans un cas, cette matrice pré-multiplie une colonne, dans l'autre elle post-multiplie une ligne.

5) « épreuve » au singulier, car il s'agit d'une abréviation de l'expression « catégorie des résultats d'épreuve », qui n'a rien à voir avec les « épreuves répétées » dont on a parfois voulu faire le fondement, sinon de la théorie des probabilités, du moins de ses applications; cfr. [II].

6) On notera que la moyenne du vecteur aléatoire $\mathbf{\mathfrak{b}}$ est un vecteur défini sans recours à une base (théorie de l'intégration dans les espaces vectoriels), de sorte que la notation $\mathbf{E}\,\mathbf{\mathfrak{b}}$ a un sens intrinsèque [il est très heureux que $(\mathbf{E}\,\mathbf{\mathfrak{b}})_p = \mathbf{E}\,(\mathbf{\mathfrak{b}}_p)$]. L'étude intrinsèque de la covariance serait un peu moins simple.

7) On dit parfois que « des variables aléatoires normales non corrélées sont indépendantes ». Cet énoncé, pris dans toute sa généralité, est faux; il est vrai pour des aléatoires (normales, nécessairement) qui sont les composantes d'une représentation (par rapport à une base certaine) d'un vecteur multinormal.

8) Plus explicitement: $\mathfrak{l}^{\star} \longrightarrow [\mathfrak{b} \longrightarrow \mathfrak{l}^{\star}\,\mathfrak{A}\,\mathfrak{b}]$.

9) « Identiquement » par rapport à la variabilité de $\mathfrak{b}$ dans B.

10) Comme on le sait, le mot « erreur » possède, en statistique, un sens très éloigné de son sens vulgaire.

11) Il s'agit là d'une variable aléatoire; la notation appropriée à ce fait est malaisée à choisir; la convention adoptée ici a, à défaut d'autre mérite, celui d'être simple.

12) Où « red » signifie « réduction » (*scil.* de la somme de carrés des erreurs).

H. Breny,
Centre interdisciplinaire d'analyse stochastique
et de recherche opérationnelle
Université de Liège.

(A suivre)