COURBE D'INFLUENCE (G8)
(26 / 08 / 2020, © Monfort, Dicostat2005, 2005-2020)

(i) La notion générale d'**influence** se réfère à une **intéraction** unilatérale, ou à une **causalité**, pouvant exister entre deux groupes de **variables**. Cette liaison peut se représenter à l'aide de la notion de **relation fonctionnelle**, dans laquelle on distingue :

(a) un groupe de **variables exogènes**, encore appelées causes, ou origines, ou stimuli, ou variables idépendantes, ou variables explicatives, ou variables contrôlées, ou variables non conditionnées, ou variables d'entrée, ou variables de droite, etc ;

(b) et un groupe de **variables endogènes**, encore appelées effets, ou destinations, ou réponses, ou variables dépendantes, ou variables expliquées, ou variables non contrôlées, ou variables conditionnées, ou variables de sortie, ou variables de gauche, etc.

(ii) Un autre concept d'**influence**, spécifique de la théorie de la **robustesse**, permet d'apprécier le rôle joué par les **observations** relatives à un **phénomène** sur une **statistique** d'intérêt.

Ce concept est distinct de celui de robustesse dans un **modèle statistique**. En effet :

(a) la **robustesse** exprime un degré de variation d'une **procédure statistique** résultant de modifications des hypothèses apportées au modèle qui fonde cette procédure. Ces hypothèses sont essentiellement relatives aux **lois de probabilité** du modèle, ie symboliquement :

Δ hypothèses (lois) → Δ procédure (statistiques) ;

(b) l'**influence** exprime un degré de variation d'une procédure pr à des variations (ou modifications) des **observations** à partir desquelles l'**inférence statistique** est réalisée à l'aide du modèle considéré, ie :

Δ observations (variables) → Δ procédure (statistiques).

La notion de **courbe d'influence** permet notamment d'apprécier ce dernier type de variation.

(iii) Soit $(\Omega, \mathscr{T}, \mathscr{P})$ un **modèle statistique** et $(\mathscr{X}, \mathscr{B}, \mathscr{P}^X)$ son image par une **va** (**échantillon**) X. On note $\mathscr{F}$ l'ensemble des **fr** F associées aux **lp** $P^X = X(P) \in \mathscr{P}^X$, et l'on fait de $\mathscr{F}$ un **espace métrique** par le choix d'une **distance** (eg **distance de PROKHOROV**).

On considère une **fonctionnelle** $t : \mathscr{F} \mapsto \mathbf{R}$ qui est $(\mathscr{B}^X, \mathscr{B}_{\mathbf{R}})$-mesurable (où $\mathscr{B}^X$ est la tribu borélienne de $\mathscr{F}$). Soit $\mathscr{E} \subset \mathscr{F}$ une **partie étoilée** en $F \in \mathscr{F}$, ie une **partie** $\mathscr{E} \in \mathscr{P}(\mathscr{F})$ tq :

(1) $[F, G] = \{F + \alpha (G - F) : \alpha \in [0, 1]\} \subset \mathscr{E}, \qquad \forall\, G \in \mathscr{E}.$

On dit que t admet une **dérivée directionnelle** (ou dérivée dans la direction $G \in \mathscr{F}$) au point $F \in \mathscr{F}$ ssi il existe une fonctionnelle $u : \mathscr{F} \times \mathscr{F} \mapsto \mathbf{R}$ tq :

(2) $\lim_{\varepsilon \to 0+} \varepsilon^{-1} . \{t (F + (\varepsilon (G - F)) - t (F)\} = u (F, G).$

On note aussi $D_G\, t (F)$ au lieu de $u (F, G)$: ce symbolisme représente ainsi la **dérivée directionnelle** de t selon G en F (ou dérivée de t dans la direction G au point F).

On dit que t admet une **dérivée dans la direction G** sur $\mathscr{E}$ ssi t admet une dérivée dans la direction G en tout point $F \in \mathscr{E}$. On note alors cette dérivée selon $u (. , G)$ ou $D_G\, t$ (dérivée de t dans la direction G).

(iv) On appelle alors **graphe d'influence (de F.R. HAMPEL)** de t au point F le graphe Γ_F de la fonction γ_F définie selon :

(3) $x \in \mathscr{X} \mapsto \gamma_F (x) = D_{\delta(x)}\, t (F) \in \mathbf{R}.$

En général, $\mathscr{X} = \mathbf{R}$ et le graphe Γ_F s'appelle alors **courbe d'influence (de F.R. HAMPEL)**. Par suite, si f est la **densité** (pr à λ_1) associée à F, on peut définir la **courbe d'influence** relative eg :

(a) à l'**espérance mathématique** $E\, X = t (F) = \int x\, dF (x)$ selon :

(4) $\gamma_F (x) = x - E\, X$;

(b) à la **médiane** (supposée unique) $Q_{1/2}\, X = t (F) = F^{-1} (1/2)$:

(5) $\gamma_F (x) = (1/2) \{\mathbf{1}_A (x) - \mathbf{1}_B (x)\} . f(0),$

où $A = \mathbf{R}_+^*$ et $B = \mathbf{R}_-^*$ (en supposant, en outre, que f est à la fois une **loi symétrique** et une **loi unimodale**, et que l'**information de FISHER** associée existe).

La définition de la courbe d'influence s'applique, en particulier, lorsque $F = F_N$ (**fonction de répartition empirique** associée au N-**échantillon aléatoire** X.

(v) Plus généralement, t peut être une **caractéristique** quelconque de F : **moment**, **quantile**, etc.

La dérivée u est souvent appelée **dérivée de R. von MISES** et elle peut s'écrire sous la forme :

(6) $u\,(F, G) = \int v_F \, dG,$

où $v_F : \mathscr{X} \mapsto \mathbf{R}$ est une fonction mesurable (qui est aussi appelée **courbe d'influence** lorsque $\mathscr{X} = \mathbf{R}$).

Si l'on choisit $G = H_x$ (**fonction de HEAVYSIDE** au point x), la fonction $u\,(F, H_x) = v_F\,(x)$ est appelée **influence** de x sur la fonctionnelle t. Si $G = F_N$ (fr empirique), la fonction :

(6) $u\,(F, F_N) = N^{-1} \sum_{n=1}^{N} v_F\,(X_n)$

représente la moyenne des influences des observations X_n .

(vi) L'influence des observations sur une fonctionnelle s'exprime aussi par la notion de **(fonction de) résistance** (cf **résistance**).