ROBUSTESSE (G8, H5, I6, J6)
(21 / 01 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **théorie de la robustesse** est une composante fondamentale de la méthodologie statistique : lorsqu'un **contexte statistique** est favorable, on préfèrera une méthode « robuste » à une méthode non robuste.

(i) L'analyse de la robustesse d'une **procédure** statistique (**sondage**, **expérience**, **classification**, **estimation**, **test**, **prévision**) se déduit des concepts de **problème de décision** et de **problème statistique**. Elle peut ainsi se concevoir sous plusieurs angles :

(a) robustesse pr à la « vraie » **famille des lois de probabilité** supposée avoir engendré des **observations** relatives à un **phénomène** donné : eg la vraie **famille** est celle des lois admettant des moments jusqu'au second ordre, mais la famille retenue (pour diverses raisons : moyens disponibles, facilités de calcul, échéances, etc) n'est que la famille des **lois normales** ;

(b) robustesse pr à la **famille des fonctions de perte** retenues (donc à celle des **fonctions de risque** associées) : eg dans un **modèle de régression multiple**, la procédure d'estimation retenue est (à tort) la **méthode des moindres carrés**, alors que la **méthode des moindres écarts absolus** serait plus appropriée, eg en raison de la présence d'observations aberrantes (cf **aberration**) ;

(c) robustesse pr aux **lois a priori** retenues par le **statisticien**, lorsque celui-ci adopte une approche bayésienne (cf **école bayésienne**, **principe bayésien**, **théorie bayésienne**) ;

(d) robustesse pr aux **observations utilisées**. Ainsi (cf supra), la présence d'observations aberrantes n'est pas sans effet sur une procédure. Cependant, dans cette situation, on parle plutôt de **« résistance »**, ou de **« sensibilité »** de la procédure aux données atypiques, ou encore d' **« influence »** de ces données sur la procédure (cf aussi **validation croisée**). La notion de robustesse est donc à distinguer de celles de **résistance** ou d'influence (cf aussi **courbe d'influence**).

Les approches précédentes ne sont pas exclusives les unes des autres. En effet, la présence d'aberrations peut tenir à l'existence d'un **mélange de lois** ou d'une **loi à queue épaisse**, ce qui relie (a) et (c). De même, le choix d'une fonction de perte, en dehors de toute considération de commodité de calcul, peut viser à atténuer l'effet des observations extrêmes sur la règle de décision à adopter, ce qui associe (b) et (c).

(ii) La théorie de la robustesse adopte souvent un cadre non paramétré, ou non paramétrique (cf **modèle paramétrique**, **paramètre**) : ceci permet notamment de définir directement, et de façon formellement « homogène », diverses **distances** entre lois, entre fonctions de perte, etc. Elle fait souvent usage de **méthodes affranchies**.

Ainsi, la robustesse peut se poser :

(a) pr à la forme analytique des lois considérées (eg les **lois normales**) ;

(b) pr à l'indépendance des observations (lorsque les données X constituent un **échantillon**) ;

(c) pr aux **caractéristiques** des lois, eg :

$(c)_1$ **caractéristiques de forme** (eg **symétrie**, **voussure**) ;

$(c)_2$ **hétéroscédasticité** dans un **modèle de régression** (cf **régression robuste**).

(iii) Le cadre de la **théorie de la décision** statistique unifie en grande partie les approches et concepts liés à la robustesse. Le **problème de la robustesse** (G.E.P. BOX, 1953) a pour objet l'étude de l'effet exercé sur une **procédure statistique** par un **jeu d'hypothèses** qui ne sont pas celles qui fondent la procédure en question. Il étudie donc en quoi le fait de s'« éloigner » d'un **problème de décision** théoriquement « correct » (mais non perçu par le statisticien ou par l'**homme de l'art**) influe sur la qualité d'une procédure. Autrement dit, et de façon générale, le problème de la robustesse analyse les **altérations des propriétés d'une procédure** lorsque les hypothèses sous-jacentes ne sont pas vérifiées ou lorsqu'elles sont modifiées.

D'un point de vue formel, ce problème s'apparente à un **problème de stabilité** : stabilité d'un **système** physique, biologique, etc.

(iv) Soit $(\Omega, \mathcal{T}, (P_\alpha)_{\alpha \in A})$ un **modèle statistique** de base, $(\mathcal{X}, \mathcal{B})$ un **espace d'observation** et $X : \Omega \mapsto \mathcal{X}$ une **variable aléatoire observable** (eg un **échantillon**). On note $(\mathcal{X}, \mathcal{B}, (P_\alpha^X)_{\alpha \in A})$ le **modèle image**, dans lequel $P_\alpha^X = X(P_\alpha)$ est la **loi** de X, $\forall \alpha \in A$. On étudie le problème de **décision statistique** $\{(\mathcal{X}, \mathcal{B}, (P_\alpha^X)_{\alpha \in A}), (D, \mathcal{B}_D), L\}$ fondé sur le modèle précédent, et dans lequel D est l'ensemble des **décisions** possibles et $L : D \times A \mapsto \mathbf{R}_+$ une **fonction de perte** donnée.

Un **problème de robustesse** se pose lorsqu'une (au moins) des hypothèse définissant le problème de décision précédent n'est pas vérifiée, c'est-à-dire (de façon extensive) lorsque :

(a) $(\mathcal{X}, \mathcal{B})$ n'est pas le « vrai » espace d'observation ;

(b) $(P_\alpha^X)_{\alpha \in A}$ n'est pas la « vraie » famille des lois qui génèrent X (ou l'observation $x = X(\omega)$) ;

(c) $(D, \mathcal{B}_D)$ n'est pas un **espace de décision** correct, donc $\Delta = D^{\mathcal{X}}$ n'est pas susceptible de contenir de « bonnes » règles de décision $\delta^\#$;

(d) L n'est pas une fonction de perte pertinente, c'est-à-dire mesurant la perte effectivement encourue par le statisticien lorsqu'il décide en fonction des états de la **Nature**.

L'espace d'observation est souvent donné (ie imposé, ou défini par le statisticien une fois pour toutes), de même que l'espace de décision. Les problèmes courants apparaissent donc surtout dans les situations (b) et (d) précédentes.

(v) **Robustesse pr à la famille des lois**. Si $\mathscr{P}_A{}^X = (P_\alpha{}^X)_{\alpha \in A}$ n'est pas la famille de lois adéquate pour représenter le phénomène aléatoire étudié, la « vraie » famille (en général inconnue du statisticien, ou non retenue par lui), notée $\mathscr{P}_B{}^X = (P_\beta{}^X)_{\beta \in B}$, peut être disjointe de la précédente, ou non. On définit une famille $\mathscr{P}^X$ comme réunion des deux familles précédentes, soit $\mathscr{P}^X = \mathscr{P}_A{}^X \cup \mathscr{P}_B{}^X$, famille qui peut se paramétrer selon $\mathscr{P}^X = (P_\theta{}^X)_{\theta \in \Theta}$, avec $\Theta = A \cup B$. On suppose cette famille identifiable (cf **identifiabilité**).

Pour étudier les répercussions d'une telle « extension » sur le problème initial, il convient d'étendre aussi la fonction de perte L : D x $A \mapsto \mathbf{R}_+$ en une fonction de perte (encore notée L) L : D x $A \mapsto \mathbf{R}_+$, avec $\Theta = A \cup B$. Si cette extension est possible (cf **extension d'une application**, **restriction d'une application**), la fonction de risque se définit de façon usuelle (cas de règles pures) :

(1) $\quad R(\delta, \theta) = E_\theta L(\delta(X), \theta), \qquad \forall \theta \in \Theta,$

où $E_\theta L(\delta(X), \theta) = \int L(\delta(x), \theta)\, dP_\theta{}^X(x)$. On note R_A (resp R_B) la restriction de la seconde application partielle de R à A (resp à B). La démarche suivie conduit à adopter une règle R_A-optimale, ie une règle de décision $\delta_A{}^\# : \mathscr{X} \mapsto D$ tq :

(2) $\quad R_A(\delta_A{}^\#, \theta) \leq R_A(\delta, \theta), \qquad \forall \delta \in \Delta, \forall \theta \in \Theta.$

En réalité, la bonne règle est une règle R_B-optimale $\delta_B{}^\# : \mathscr{X} \mapsto D$, définie de façon analogue. La règle effectivement retenue $\delta_A{}^\#$ est dite **robuste** pr à la règle $\delta_B{}^\#$ ssi le fait de passer de la famille $\mathscr{P}_A{}^X$ à la famille $\mathscr{P}_B{}^X$ n'influe guère sur $\delta_A{}^\#$, ou ssi la loi de probabilité de $\delta_B{}^\#$ est peu différente de celle de $\delta_A{}^\#$ (elle varie donc peu lorsque la famille $\mathscr{P}_A{}^X$ est altérée). Plus précisément, si d est une **distance** définie sur $\mathscr{P}^X$ (cf **espace métrique**) et si m est une **métrique** définie sur $\Delta = D^{\mathscr{X}}$, $\delta_A{}^\#$ est **robuste** pr à $\delta_B{}^\#$ ssi la variation :

(3) $\quad v(\delta, \alpha, \beta) = |R(\delta, \beta) - R(\delta, \alpha)| = |R_B(\delta, \beta) - R_A(\delta, \alpha)|$

du risque associé à toute règle de décision $\delta \in \Delta$ est « négligeable » pr à celle d $(P_\beta{}^X, P_\alpha{}^X)$ des lois. Ceci entraîne que la variation m $(\delta_B{}^\#, \delta_A{}^\#)$ est, elle aussi, négligeable devant d $(P_\beta{}^X, P_\alpha{}^X)$.

Autrement dit, $\delta_A^{\#}$ est **robuste** pr à $\delta_B^{\#}$ ssi, pour tout $\varepsilon > 0$, il existe $\eta > 0$ tq :

(4) $\sup d(P_\beta^X, P_\alpha^X) \leq \eta \quad \Rightarrow \quad \sup v(\delta, \alpha, \beta) \leq \varepsilon, \qquad \forall \delta \in \Delta,$

les opérations « sup » étant prises sous les contraintes resp $(P_\beta^X, P_\alpha^X) \in (\mathscr{P}_\beta^X \times \mathscr{P}_\alpha^X)$ et $(\alpha, \beta) \in A \times B$.

La notion de **robustesse locale** se déduit de (4) en y remplaçant $\mathscr{P}^X$ par une de ses **parties** strictes $\mathscr{P}_0^X$ contenant $\mathscr{P}_A^X$, ou en remplaçant Θ par une de ses parties strictes Θ_0 contenant A.

(vi) **Robustesse pr à la fonction de perte**. Si L n'est pas la fonction de perte adéquate, notée L_0 , le problème peut se poser comme suit. On suppose définie, sur l'ensemble $\mathscr{L}$ des fonctions de perte $L : D \times \Theta \mapsto \mathbf{R}_+$ possibles, une distance e, et l'on étudie comment varient les règles de décision optimales lorsque la fonction de perte utilisée diffère de L_0 . En notant alors R le risque associé à L, $\delta^{\#}$ une règle de décision R-optimale :

(5) $R(\delta^{\#}, \theta) = \inf_{\delta \in \Delta} R(\delta, \theta),$

et $\delta_0^{\#}$ la règle optimale au sens du risque R_0 associé à L_0 (et définie de façon analogue à $\delta^{\#}$), on dit que la règle de décision $\delta^{\#}$ effectivement utilisée est **robuste** pr à la règle adéquate $\delta_0^{\#} \in \Delta$ ssi le fait de passer de la fonction L_0 à la fonction L influe peu sur $\delta_0^{\#}$, ou ssi la lp de $\delta^{\#}$ est peu différente de celle de $\delta_0^{\#}$: elle varie alors peu lorsque la perte utilisée $L \in \mathscr{L}$ n'est pas la fonction correcte $L_0 \in \mathscr{L}$.

Plus précisément, $\delta^{\#}$ est **robuste** pr à $\delta_0^{\#}$ ssi la variation :

(6) $v(\delta, \theta) = |R(\delta, \theta) - R_0(\delta, \theta)|$

du risque associé à toute règle de décision $\delta \in \Delta$ est négligeable pr à celle e (L, L_0) des fonctions de perte, ce qui entraîne alors que la variation m $(\delta_2^{\#}, \delta_1^{\#})$ des règles de décision optimales est, elle aussi, négligeable devant e (L, L_0).

(vii) Les deux notions précédentes de robustesse sont les plus courantes. Elles expriment donc la faible altération de la lp d'une règle optimale robuste lorsque la lp (resp la fonction de perte) du modèle est perturbée. Si à chaque loi $P^X \in \mathscr{P}^X$ (resp à chaque perte $L \in \mathscr{L}$) on associe une règle optimale $\delta^{\#}$, et si une telle règle admet pour lp la loi $P^{\delta\#}$, on peut munir la famille $\mathscr{P}^{\delta\#}$ de ces lois d'une distance h (où δ# désigne $\delta^{\#}$). Par suite, la correspondance $P^X \mapsto P^{\delta\#}$ (resp $L \mapsto P^{\delta\#}$) est continue relativement aux mesures d et h (resp e et h) ssi la règle optimale de loi $P^{\delta\#}$ est robuste.

La notion de robustesse précédente est une notion de **robustesse globale**.

On lui substitue souvent une notion de **robustesse locale**, qui exprime la continuité des correspondances précédentes seulement sur un voisinage $\mathscr{P}_0^X$ de $\mathscr{P}_1^X$ (resp un voisinage $\mathscr{L}_0$ de L_0).

Les fonctions d, e, h ou m ne sont pas toujours de vraies distances : en pratique, elles peuvent être des **semi-distances** ou des pseudo-distances, ie des fonctions ne vérifiant pas tous les axiomes des distances.

L'analyse abstraite des problèmes de robustesse fait aussi appel à des topologies plus générales que celles des **espaces métriques**, définies par les distances précédentes.

Parmi les distances les plus utilisées, on peut citer : la **distance de BHATTACHARYYA**, la **distance de CRAMER-MISES**, la **distance de HELLINGER**, la **distance de KOLMOGOROV**, la **distance de LÉVY** ou encore la **distance de PROKHOROV**.

Enfin, les notions de **dérivée de GATEAU** ou de **dérivée de GATEAU-VOLTERRA** sont souvent adaptées aux problèmes d'analyse fonctionnelle qui se posent en théorie de la robustesse.

(viii) Cette théorie est donc centrée sur la **stabilité des décisions (optimales)** en fonction du choix d'un **modèle** ou, plus généralement, d'un problème statistique. Dans ce dernier cadre, on traite un problème $\{(\mathscr{X}_1 , \mathscr{B}_1 , \mathscr{P}_1^X), (D_1 , \mathscr{D}_1) , \mathscr{L}_1\}$ alors que le vrai problème à traiter est $\{(\mathscr{X}_2 , \mathscr{B}_2 , \mathscr{P}_2^X), (D_2 , \mathscr{D}_2), \mathscr{L}_2\}$, où les $\mathscr{L}_i$ (i = 1, 2) sont deux familles de fonctions de perte. Ces deux problèmes peuvent être « plongés » (cf **plongement**) dans un problème de décision unique global $\{(\mathscr{X}, \mathscr{B}, \mathscr{P}^X), (D, \mathscr{D}), \mathscr{L}\}$, dans lequel $\mathscr{X} = \mathscr{X}_1 \cup \mathscr{X}_2$, $\mathscr{B} = \sigma(\mathscr{B}_1 \cup \mathscr{B}_2)$ (**tribu engendrée** par la réunion des tribus), $\mathscr{P}^X = \mathscr{P}_1^X \cup \mathscr{P}_2^X$ est la famille des lois de probabilité P^X définies selon :

(7) $$P^X \in \mathscr{P}^X \quad\Leftrightarrow\quad \{\exists\, i \in \{1, 2\} \text{ et } P_i^X \in \mathscr{P}_i^X \text{ tq } P^X{}_{/\mathscr{B}(i)} = P_i^X \text{ et } P^X{}_{/(\mathscr{B} \setminus \mathscr{B}(i))} = 0\},$$

où $P^X{}_{/\mathscr{C}}$ désigne la restriction de P^X à $\mathscr{C}$, et où $\mathscr{B}(i)$ désigne $\mathscr{B}_i$ (i = 1, 2). Autrement dit, $\mathscr{P}^X$ est une extension de $\mathscr{P}_1^X \cup \mathscr{P}_2^X$ à $\mathscr{B}$. Enfin, on note $D = D_1 \cup D_2$ et l'on désigne par $\mathscr{L}$ la famille des fonctions de perte L définies par :

(8) $$L \in \mathscr{L} \quad\Leftrightarrow\quad L = \mathbf{1}(D_1) \,.\, L_1 + \mathbf{1}(D_2) \,.\, L_2 ,$$

où **1**(A) dénote la **fonction indicatrice** d'une partie A.

En général, si les deux problèmes de décision ne sont pas trop « différents », on a $\mathscr{X}_1 \cap \mathscr{X}_2 \neq \varnothing$ et $D_1 \cap D_2 \neq \varnothing$. On munit alors $\mathscr{P}^X$ d'une distance d, $\mathscr{L}$ d'une distance e, l'ensemble $\mathscr{P}^{\#}$ des lois $P^{\delta\#}$ d'une distance h et l'ensemble $\Delta = D^{\mathscr{X}}$ des règles de décision d'une métrique m.

(ix) Lorsque X est un échantillon, on peut distinguer les notions de :

(a) **robustesse à distance finie**. Dans ce cas, la taille d'échantillon(nage) est donnée : X est un échantillon de taille N finie (avec, parfois, $N << \infty$), et l'on étudie diverses procédures à N fixé ;

(b) **robustesse asymptotique**. Dans ce cas, la taille N d'échantillon(nage) est « importante » ($N >> 0$, d'où l'assimilation avec $N \to \infty$) et les loi de probabilité concernées sont les **lois asymptotiques** suivies par des **statistiques** d'intérêt. Ce sont les variations des lois asymptotiques $P^{\delta(N)\#} = \mathscr{L}(\delta_N^{\#})$, ainsi que la stabilité des procédures pr à ces variations, qui sont analysées en fonction des hypothèses. Cette approche bénéficie souvent de la simplicité des calculs effectués à partir des lois limites.

(x) La notion de robustesse est aussi liée à celle de **spécification** de modèle. En effet, une **erreur de spécification** peut s'interpréter comme le choix d'un modèle erroné, mais qui peut être acceptable si le problème de décision statistique est robuste. Ceci nécessite donc une définition appropriée de la notion de **test de robustesse** : de ce point de vue, une procédure robuste minimise l'inconvénient dû à l'**erreur de spécification** consistant à utiliser un modèle au lieu d'un autre. Le modèle retenu par le statisticien est souvent plus « simple » (eg un modèle paramétrique parcimonieux) que le « vrai » modèle, lequel peut être partiellement connu ou même inconnu (cf **parcimonie**). Le problème de la spécification d'un modèle doit donc porter à la fois sur la recherche du modèle le plus adéquat et sur un test de cette adéquation.

(xi) Les définitions précédentes s'étendent aux règles de décision aléatoires (ou mixtes).

De même, dans un cadre bayésien, on peut étudier deux types de robustesse :

(a) **robustesse de la loi a posteriori**. Le raisonnement est analogue à ce qui précède (dépendance de cette loi pr aux hypothèses) ;

(b) **robustesse de la règle de décision bayésienne** pr à la loi a priori elle-même. Ici, Θ est muni d'une **tribu de parties** $\mathscr{B}_\Theta$ sur laquelle la **loi a priori** Π peut subir des altérations (eg une contamination : cf **contamination des lois**). En effet, il importe de savoir si, lorsque la loi a priori n'est pas correcte (eg erreur de jugement), la règle de décision qui en résulte (règle de BAYES) n'est pas trop modifiée ou conserve encore approximativement des propriétés d'**optimalité**.