

Astérisque

DIDIER DACUNA-CASTELLE

CATHERINE HUBER

CLAUDE DENIAU

GEORGES OPPENHEIM

CLAUDE VIANO

ROBERT AZENCOTT

CATHERINE HUBER

LUCIEN BIRGÉ

VALENTINE COSTA

JEAN DESHAYES

PHILIPPE JOLIVALDT

PHILIPPE TRÉCOURT

DOMINIQUE PICARD

Théorie de la robustesse et estimation d'un paramètre

Astérisque, tome 43-44 (1977)

http://www.numdam.org/item?id=AST_1977__43-44__1_0

© Société mathématique de France, 1977, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

TABLE DES MATIÈRES

	Pages
Présentation du Séminaire	3
CHAPITRE I	
D.DACUNHA-CASTELLE - Modèles et robustesse	5
CHAPITRE II	
C.HUBER - Exemples d'instabilité et voisinages d'un modèle	11
CHAPITRE III	
Définitions et rappels	23
CHAPITRE IV	
C.DENIAU-G.OPPENHEIM-C.VIANO - M-estimateurs (Maximum de vraisemblance modifié)	31
CHAPITRE V	
R.AZENCOTT - Estimation d'un paramètre de translation à partir de tests de rang	41
CHAPITRE VI	
C.HUBER - L-estimateurs (Combinaisons linéaires des sta- tistiques d'ordre)	65
CHAPITRE VII	
L.BIRGÉ - Tests minimax robustes	87
CHAPITRE VIII	
C.DENIAU-G.OPPENHEIM-C.VIANO - Exemples de tests minimax- robustes. Intervalles de confiance robustes	135
CHAPITRE IX	
C.DENIAU-G.OPPENHEIM-C.VIANO - Robustesse : π -robustesse et minimax-robustesse	151
CHAPITRE X	
C.DENIAU-G.OPPENHEIM-C.VIANO - Robustesse des M-estima- teurs	167
CHAPITRE XI	
R.AZENCOTT - Robustesse des R-estimateurs	189

CHAPITRE XII	
C.HUBER - Robustesse des L-estimateurs	203
CHAPITRE XIII	
V.COSTA-J.DESHAYES - Comparaison des R-L-M estimateurs	209
CHAPITRE XIV	
C.DENIAU-G.OPPENHEIM-C.VIANO - Courbe d'influence et sensibilité	239
CHAPITRE XV	
P.JOLIVALDT-P.TRÉCOURT - Estimateurs d'un paramètre de position : aspects numériques	253
CHAPITRE XVI	
D.PICARD - Estimateurs adaptatifs dans un modèle de ré- gression	275
Bibliographie	295

Présentation du Séminaire.

Le projet initial était de faire le point sur la théorie de la robustesse à partir d'une étude de la littérature existant sur ce sujet. En fait nous nous sommes limités pour l'essentiel aux deux problèmes les plus étudiés (et de loin) : l'estimation d'un paramètre de position et la théorie des tests. Les autres problèmes abordés dans le cadre de la théorie de la robustesse (modèle linéaire et régression par exemple) ne seront pas, sauf exception, mentionnés ici. Les idées, liées à la théorie de la robustesse concernant ces sujets, sont, en effet, identiques, à celles introduites à propos de deux problèmes cités. Cependant, ces problèmes ont amené, dans la pratique, des techniques assez fines et utiles [2], [49] . Conçu pour faire le point sur une théorie, l'exposé ne prétendait pas apporter des idées originales. Cependant, nous nous sommes attachés à donner des démonstrations de nombreux théorèmes, souvent énoncés, et non démontrés à notre connaissance, ainsi qu'à rectifier certaines démonstrations erronées. De ce point de vue, nous espérons fournir un outil aux statisticiens.

Le plan du travail est le suivant :

Les chapitres I, II, III sont une introduction à la robustesse et donnent les définitions et notations utilisées.

Les chapitres IV à VI traitent des classes usuelles d'estimateurs de paramètres de position (classes M, R, L). L'essentiel consiste dans les études de normalité asymptotique, certaines démonstrations (notamment L-estimateurs) améliorent les méthodes

déjà utilisées.

Les chapitres VII et VIII traitent de la théorie des tests. On trouvera en particulier une généralisation et une rectification des résultats connus de HUBER et STRASSEN.

Les exposés IV à XIII, traitent de l'estimation robuste d'un paramètre de position. Les démonstrations sont, à notre connaissance les premières à être publiées.

Les exposés XIV et XV, traitent des aspects numériques et descriptifs à partir du livre de ANDREWS et Co [2].

Enfin, nous avons conclu par l'exposé XVI sur les méthodes adaptatives qui nous semblent s'intégrer comme une suite logique à certains des problèmes abordés.

[Certains exposés, à caractère essentiellement pédagogique, touchant notamment la contiguïté, n'ont pas été publiés ici] .

CHAPITRE I
MODÈLES ET ROBUSTESSE
par
Didier DACUNHA-CASTELLE

Dans tout problème statistique se pose (ou devrait se poser) le choix d'un modèle.

Par modèle statistique nous entendons un ensemble E de valeurs observables, une σ -algèbre \mathfrak{B} , et un ensemble \mathfrak{M} de probabilités sur (E, \mathfrak{B}) . L'analyse, avant l'expérience, du phénomène aléatoire étudié, conduit au choix de \mathfrak{M} dans l'ensemble des lois de probabilité sur (E, \mathfrak{B}) . Une distinction classique (métamathématique) entre modèles est celle de modèle paramétrique et non paramétrique. La physique par exemple conduit le plus souvent à des situations où \mathfrak{M} est un ensemble naturellement indexé (terme non défini mathématiquement) par un paramètre $\theta \in \mathbb{R}^k$. A une certaine valeur du paramètre correspond la probabilité qui régit le phénomène étudié. A l'opposé, la biologie, la psychologie, et d'autres situations sur l'analyse théorique préexpérimentale et encore peu mathématisée, amènent au choix d'un ensemble \mathfrak{M} très vaste, sans paramétrage naturel. Cependant cette distinction, historiquement importante, tend à s'estomper. Il reste pour éviter une confusion fréquente, à dire que la théorie non paramétrique, n'est pas la théorie de la robustesse.

Cette dernière est née de problèmes très concrets posés à partir de la statistique

paramétrique et pour l'essentiel à partir de l'utilisation de modèles gaussiens ([3], [22], [80]).

Il y a toujours distorsion entre un phénomène, sa modélisation (ici mathématique) avant l'expérience et le modèle (ici statistique) qui sert à exploiter les résultats de l'expérience. Examinons ce dernier point plus en détail. L'expérimentateur sait, a priori, que dans les mesures interviennent des erreurs d'expérimentation dues à la sensibilité intrinsèque du matériel, mais aussi des erreurs de type numérique (cumul d'erreurs d'arrondis lorsque l'on travaille à une certaine échelle sur ordinateur), erreurs grossières de plusieurs types (mesures aberrantes, erreurs de transcription, etc...) dont l'importance est grande dans beaucoup de domaines même voisins de la physique expérimentale. On sait qu'il y a différents types qualitatifs de mesures allant de mesures très précises de la physique (qui sont employées sans inconvénient dans le modèle initial) jusqu'à des mesures très imprécises ou à des simples classements des observations. Il est donc des cas où il y a contradiction entre un modèle gaussien préexpérimental et le résultat des expériences. En agronomie c'est le cas assez fréquent d'une parcelle donnant un rendement aberrant, c'est-à-dire dont l'écart à une moyenne estimée est très (terme de la pratique) supérieur à l'écart estimé dans le modèle gaussien d'analyse de la variance. Une pratique fréquente est de se fixer une règle "très souple" [3] pour définir ces données aberrantes, d'examiner de nouveau après l'expérience les conditions propres à la parcelle en question, d'écarter dans un premier temps cette parcelle, puis d'y revenir pour voir si l'effet aberrant ne résultait pas en fait d'un phénomène autre que celui étudié. Un exemple très différent est celui de l'économie et de la sociologie. Supposons avoir à étudier des séries chronologiques très irrégulières, par exemple celles des prix de matières premières ou celles des conflits du travail par secteur. La seule manière opérationnelle de modéliser ces séries est d'utiliser un processus à variance finie (ce qui est probablement très injustifié). La réduction de la variance et la recherche d'un modèle d'analyse conduira presque toujours à exclure les années de crise et d'effondrement des prix pour les premiers cas, les années de grands conflits type mai 68 pour le

deuxième. Cette exclusion est évidemment des plus arbitraire. Elle consiste à considérer le phénomène comme représentant tantôt le phénomène régulier modélisable et tantôt un phénomène aberrant. La même remarque vaut en économie pour calculer un coefficient saisonnier, les mois trop irréguliers sont écartés, mais la pratique justifie amplement ces procédures.

En résumé, il existe une série de pratiques consistant à rejeter, au moins dans un premier temps, les données qui rendraient le modèle trop "invraisemblable". La règle de rejet est heuristique ; dans les meilleurs cas, les données rejetées sont examinées soigneusement (puisque en fait l'examen soigneux des données aberrantes par rapport au modèle préexpérimental est la source de tout progrès scientifique). Cet examen est nécessaire dans tous les cas (qui nous intéressent essentiellement ici) où l'origine possible des aberrations est mal déterminée.

La situation économique est typique du cas ordinaire où le modèle de fonctionnement "normal" est occulté par un modèle parasite, représentant ici les crises exceptionnelles. Des situations analogues peuvent se trouver en biologie ou en médecine par suite du mélange de population ou de symptômes, et dans certains cas, on ne peut attribuer une origine aux données aberrantes, erreurs grossières ou expérience parasitée. Et la distinction peut souvent être d'un intérêt secondaire.

Dans la démarche purement statistique (test, estimation), la coutume est de se rassurer, en vérifiant si le traitement mathématique que l'on fait sur des données provenant d'un modèle que l'on soupçonne assez différent de celui que l'on a posé, reste valable. Ceci bien sûr n'est pas propre à la statistique, mais a ici une dimension spéciale. Si la réponse est positive, on dit avoir affaire à une procédure robuste. Ces études de robustesse sont classiques en analyse de la variance (hypothèse d'égalité des variances des traitements, hypothèse d'indépendance des observations, etc... [80]).

Comme nous l'avons souligné, les erreurs grossières, les erreurs d'arrondis, les phénomènes parasites se mélangent dans l'expérience de manières si diverses qu'elles excluent une théorie générale du problème de la robustesse qui les examinerait une à

une. Aussi la théorie de la robustesse que nous allons étudier est une théorie mathématique qui intègre d'emblée tous ces éléments en une notion de déviation entre le modèle préexpérimental et le modèle qui sert à exploiter l'expérience. Cette déviation est quantifiée, en remplaçant le modèle initial \mathcal{M} , par un modèle étendu \mathcal{M}' , $\mathcal{M}' \supset \mathcal{M}$, de manière à ce que si la probabilité P représentant le phénomène avant expérience est dans \mathcal{M} , la probabilité P' décrivant le phénomène sous expérience est dans \mathcal{M}' . Insistons donc sur le fait que le choix d'un modèle étendu (modèle conduisant à des procédures statistiques robustes) se fait sur deux types de connaissances a priori :

- 1) celle du modèle "théorique" (loi en physique) \mathcal{M} du phénomène étudié,
- 2) celles des conditions expérimentales plausibles qui conduisent à étendre le modèle de \mathcal{M} à \mathcal{M}' .

(On ne cachera pas qu'il y aurait une optique bayésienne des deux connaissances que nous n'étudierons presque pas).

La théorie de la robustesse devrait donc (et c'est le cas souvent) aboutir à des conclusions justifiant les pratiques usuelles. L'essentiel étant de remarquer que les procédures habituelles (optimales ou non) utilisées sur des problèmes paramétriques sont instables dès que l'on s'éloigne du modèle. Il ressort de cette analyse que l'on aura besoin de définir sur les modèles quelque chose ressemblant à une topologie, de manière à rechercher des procédures stables au voisinage des modèles préexpérimentaux, le voisinage étant justement le modèle d'exploitation de l'expérience.

Un bilan autre que mathématique de la théorie de la robustesse est difficile à faire. Il semble qu'elle ait eu, ces dernières années le mérite essentiel de faire s'interroger un nombre important de gens sur la validité de la méthodologie scientifique qu'ils utilisaient. Il ne semble pas pour autant qu'elle ait conduit à un grand perfectionnement des méthodes utilisables. Cependant le champ d'application de la théorie s'élargit. Partie des problèmes propres au modèle gaussien, elle aborde maintenant ceux de la statistique non paramétrique (en regardant par exemple l'hypothèse d'indépendance

des observations, de la statistique descriptive (en regardant les problèmes de stabilité des analyses de données), de la statistique des processus.

Elle est donc une partie vivante de la statistique mathématique.

Didier DACUNHA-CASTELLE

Mathématiques
Université Paris-Sud
91405 - ORSAY

CHAPITRE II

EXEMPLES D'INSTABILITÉ ET VOISINAGES D'UN MODÈLE

par

Catherine HUBER

Dans la première partie, nous verrons quelques exemples illustrant le mauvais comportement des procédures usuelles en statistique paramétrique dès qu'on s'écarte un peu du modèle. Ce modèle sera le plus souvent gaussien, et les déviations par rapport au modèle du type contamination (2 a), ce qui permet un calcul explicite d'évaluation et de comparaison de performances.

Dans la deuxième partie, nous verrons quelques types de modèles étendus permettant le mieux de tenir compte des déviations attendues par rapport au modèle initial. Ce sont, en général, des voisinages du modèle initial pour des distances convenablement choisies sur l'ensemble des probabilités.

1 - EXEMPLES D'INSTABILITÉ.

1a Tests : Considérons le problème de test le plus simple :

P_0 et P_1 étant deux probabilités complètement spécifiées de densités respectives p_0 et p_1 par rapport à une mesure σ finie μ et (X_1, \dots, X_n) étant un n-échantillon d'une loi F , il s'agit de tester H_0 ($F = P_0$) contre H_1 ($F = P_1$).

Le lemme de Neyman et Pearson donne alors les tests les plus puissants de leur niveau

qui sont fondés sur la statistique

$$T = \prod_{i=1}^n \frac{p_1(X_i)}{p_0(X_i)}.$$

Il est bien clair que T est extrêmement sensible à la présence de valeurs aberrantes qui peuvent introduire un facteur très voisin de 0 ou de $+\infty$ dans le produit qui définit T . Dans la pratique, depuis longtemps, on a couramment "éliminé" par des procédures souples et souvent non justifiées mathématiquement, les valeurs observées qui donnaient lieu à des valeurs extrêmes et apparemment aberrantes de $\frac{p_1}{p_0}$, ce qui revient à tronquer le rapport de vraisemblance $\frac{p_1}{p_0}$.

Si maintenant on sait, a priori, que la conduite de l'expérience est telle qu'en réalité on teste H'_0 ($F \in \mathcal{H}_0$) contre H'_1 ($F \in \mathcal{H}_1$) où \mathcal{H}_0 (respectivement \mathcal{H}_1) est un ensemble de probabilités englobant P_0 (respectivement P_1) et toutes les déviations auxquelles peut donner lieu l'expérience, le critère d'optimalité peut être de type minimax :

pour un niveau donné sur \mathcal{H}_0 , maximiser le minimum de la puissance sur \mathcal{H}_1 .

Il se trouve que pour certains types de modèles étendus $(\mathcal{H}_0, \mathcal{H}_1)$, il existe dans $\mathcal{H}_0 \times \mathcal{H}_1$ un couple "minimal" (Q_0, Q_1) [VII, 45, 46] qui a la propriété suivante :

les tests optimaux de Q_0 contre Q_1 (donnés par le lemme de Neyman et Pearson) sont minimax pour tester \mathcal{H}_0 contre \mathcal{H}_1 . Et $\frac{q_1}{q_0}$ est une version tronquée du rapport $\frac{p_1}{p_0}$.

Après ces remarques générales concernant l'utilisation pratique du lemme de Neyman et Pearson, considérons un exemple particulier qui illustre la grande variabilité des performances d'un test paramétrique très courant (le test de Student) comparées à celles d'un test non paramétrique d'utilisation très simple (le test de Wilcoxon) :

le problème est de tester, à partir d'un n -échantillon X_1, \dots, X_n , l'hypothèse H_0 ($\theta = 0$) contre H_1 ($\theta > 0$) sur le modèle $\{F(x-\theta), \theta \in \mathbb{R}, F \in \Lambda\}$. Si Λ est réduit aux seules lois $N(0, \sigma^2)$, $\sigma^2 \in \mathbb{R}_*^+$, on sait que le test de Student est uniformément le plus puissant dans

dans la classe des tests sans biais ; sa région critique est du type

$$R = \{\bar{X}_n > C s_n\} \quad \text{où} \quad s_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

et C est une constante positive.

Si Λ est l'ensemble des lois continues et symétriques par rapport à 0, on peut utiliser le test de Wilcoxon [31], fondé sur la statistique $W_n = \sum_{\{i | X_i > 0\}} R_i$, où R_i est le rang de $|X_i|$ dans la suite des $|X_j|$, $j = 1, \dots, n$, ordonnés par ordre croissant, et de région critique du type

$$R = \{W_n > C'\}, \quad C' \text{ constante positive.}$$

On peut calculer l'efficacité asymptotique relative des deux tests [62] et on obtient

$$e_{W/T} = 12 \sigma^2 \left[\int_{\mathbb{R}} f^2(x) dx \right]^2$$

où σ^2 est la variance de la loi F et f sa densité par rapport à la mesure de Lebesgue.

Cette efficacité varie entre 0,864, sa valeur minimum, obtenue lorsque F est à support compact (un morceau de parabole : $f(x) = b(a^2 - x^2)1_{[-a+a]}(x)$, $a = \sqrt{5}$, $b = \frac{3}{20\sqrt{5}}$), et $+\infty$, cette dernière valeur étant obtenue dans de nombreux cas usuels (pas de moment d'ordre deux pour F , $\int_{\mathbb{R}} f^2(x) dx$ non convergente).

Pour la loi normale, $e_{W/T}$ vaut 0,955, c'est-à-dire qu'il faudra en gros 5 % d'observations supplémentaires pour obtenir, avec le test de Wilcoxon, la même puissance qu'avec le test de Student. Ce résultat a l'inconvénient d'être asymptotique, mais des études de Monte Carlo [39,61,63] ont montré qu'il était valable pour de très petits échantillons : par exemple, pour $5 \leq n \leq 10$ et le niveau α voisin de 0,05.

Or le test de Wilcoxon n'est pas parmi les tests non paramétriques, celui qui a la meilleure performance lorsque le modèle est normal [31] ; cependant la grande stabilité de ses performances comparées à celles du test de Student permettent de penser qu'il est avantageux, en la présence de déviations par rapport au modèle gaussien, de l'employer plutôt que celui de Student. Dans le cas où le modèle étendu par rapport au modèle gaussien est suffisamment "petit", on peut développer une

procédure adéquate, dite robuste, dont nous verrons des exemples. Les tests non paramétriques correspondent à des modèles étendus très vastes.

1b Estimateurs :

1b1. Estimation d'un paramètre de position :

Il est tentant, quoique pas toujours justifié, de modéliser les erreurs grossières par une loi de grande variance, notamment dans le cas gaussien. Par exemple, si

l'on veut estimer la moyenne θ d'une loi $N(\theta, \sigma^2)$, σ^2 connu, on peut, pour tenir compte des mesures aberrantes, choisir le modèle suivant dit normal contaminé

$$(1.1) \quad \mathcal{H} = \{(1-\epsilon)N(\theta, \sigma^2) + \epsilon N(\theta, k\sigma^2), k > 1, \epsilon > 0, \sigma^2 > 0 \text{ constantes fixées}, \theta \in \mathbb{R}\}.$$

Ce modèle est utilisable aussi pour un mélange de populations dont l'une est supposée assez rare (dans \mathcal{H} , ϵ est toujours supposé petit) et de grande variabilité. Il est difficile de déceler, même sur un assez grand échantillon, une telle contamination, comme le montrent les indications suivantes :

Considérons un échantillon de 1000 observations de X , dont la loi est du type précédent (1.1) avec un coefficient de contamination $\epsilon = 0,01$. Ce coefficient, bien que très faible, est suffisant (voir le tableau ci-dessous) pour modifier sensiblement l'efficacité de la moyenne empirique. Il y aura en moyenne 10 observations provenant de la loi contaminante, plus étalée, mais, parmi ces observations, seules donneront l'éveil celles qui, se situent en dehors de la zone $[-2,5\sigma, 2,5\sigma]$ où les fonctions de répartition des lois contaminées et non contaminées coïncident pratiquement. Or il y en aura en moyenne 40 % : si X a la loi $N(0, 9\sigma^2)$ et si φ désigne la fonction de répartition de la loi $N(0, 1)$, $P(|X| > 2,5\sigma) = 2\varphi(-0,83) \neq 0,4$, c'est-à-dire que l'on peut s'attendre à avoir, en moyenne, deux observations à droite et deux autres à gauche de l'intervalle $[-2,5\sigma, 2,5\sigma]$ et provenant de la loi contaminante. Mais il y aura aussi, hors de cet intervalle, en moyenne, 6 observations à gauche et 6 autres à droite provenant du constituant principal, puisque, si X a la loi $N(0, \sigma^2)$, $P(|X| > 2,5\sigma) = 12 \cdot 10^{-3}$.

Donc, à moins qu'une (ou plusieurs) des 4 valeurs provenant de $N(0, \sigma^2)$ ne soit

très grande, il ne sera guère possible de détecter que le modèle gaussien ne convient pas.

Si cela arrive, on a en général tendance à éliminer la (ou les) valeurs aberrantes et à considérer que le reste est gaussien, procédure dont la légitimité dépend évidemment du contexte (erreurs grossières ou mélange de populations) mais qui est souvent mauvaise et mal justifiée par des tests peu efficaces [3], [30].

Pour aller plus loin, supposons que la loi de l'expérience soit $(1-\epsilon)N(\theta, 1) + \epsilon N(\theta, 9)$. La moyenne empirique

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

est un estimateur sans biais de θ , son efficacité asymptotique est la limite, quand n tend vers l'infini, de $\frac{1}{I_n(\epsilon, \theta) V_n(\epsilon)}$ où $V_n(\epsilon) = \text{Var}(\bar{X}_n) = \frac{1+8\epsilon}{n}$ et $I_n(\epsilon, \theta)$ est

l'information de Fisher. Un calcul numérique donne les résultats suivants

Fraction ϵ de contamination	0,00	0,02	0,05	0,10
Efficacité asymptotique de la moyenne empirique	1,00	0,90	0,80	0,70.

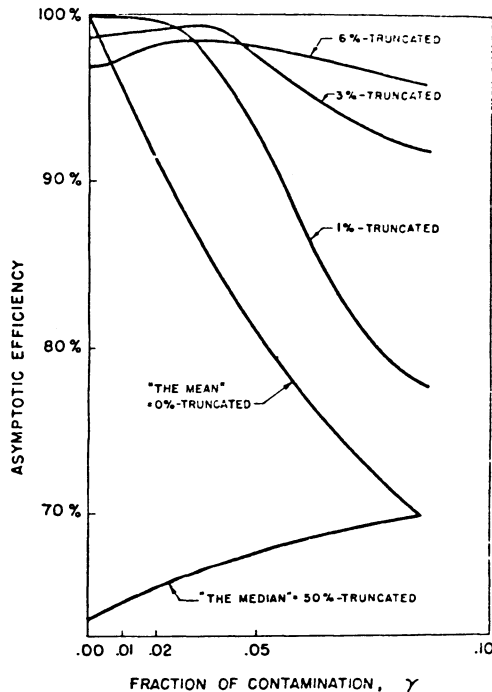
Si l'on sait que le modèle est contaminé, et si l'on connaît le degré de contamination (ce qui est évidemment une situation concrète très particulière), on peut trouver un estimateur de θ asymptotiquement efficace, soit en utilisant la méthode du maximum de vraisemblance (mais pour la loi contaminée $(1-\epsilon)N(\theta, 1) + \epsilon N(\theta, 9)$, cela donne lieu à des calculs compliqués), soit en prenant la combinaison linéaire optimale des statistiques d'ordre [VI, paragraphe 4].

Une modification très simple de la moyenne empirique consiste à la remplacer par une moyenne α tronquée : si α est dans $]0, \frac{1}{2}[$, on appelle moyenne α -tronquée de l'échantillon (X_1, \dots, X_n) de statistiques d'ordre $(X_{(1)}, \dots, X_{(n)})$ la statistique

$$\bar{X}_\alpha = \frac{1}{n-2\llbracket n\alpha \rrbracket} [X_{\llbracket n\alpha \rrbracket+1} + X_{\llbracket n\alpha \rrbracket+2} + \dots + X_{n-\llbracket n\alpha \rrbracket}]$$

où $\llbracket n\alpha \rrbracket$ désigne la partie entière de $n\alpha$. C'est la moyenne des observations obtenues en éliminant les $\llbracket n\alpha \rrbracket$ plus petites et les $\llbracket n\alpha \rrbracket$ plus grandes.

Les moyennes α tronquées contiennent la moyenne empirique, obtenue pour $\alpha=0$, et la médiane est un cas limite obtenu lorsque α tend vers $\frac{1}{2}$. La figure suivante illustre le comportement de l'efficacité asymptotique des moyennes tronquées en fonction du degré ϵ de contamination.



Efficacité asymptotique des moyennes tronquées en fonction du degré ϵ de contamination. (Tukey, [93]).

On voit sur la figure précédente que l'efficacité asymptotique de \bar{X} décroît très vite, alors qu'il suffit de tronquer à 1 % pour obtenir, au prix d'une perte minime d'efficacité pour le modèle strictement normal (0,4 %), un gain qui peut dépasser 12 %. On a même un résultat plus fort : tout estimateur optimal pour une quelconque valeur de ϵ comprise entre 0,01 et 0,10 a une efficacité asymptotique supérieure à 0,96 pour toute la zone de contamination $\epsilon \in [0; 0,10]$. L'un quelconque de ces estimateurs est donc préférable à \bar{X} qui voit son efficacité décroître régulièrement de 1 à 0,7

dans cette même zone (tableau et Fig.).

1b 2. Estimation d'un paramètre d'échelle.

Pour le modèle gaussien $(N(0, \sigma^2))^{\otimes n}$, considérons les deux estimateurs sans biais de σ fondés sur les statistiques

$$S_n = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}} \quad \text{et} \quad S_n^* = \frac{\sum_{i=1}^n |X_i|}{n} ;$$

Comme $E_\sigma(S_n)$ et $E_\sigma(S_n^*)$ sont proportionnelles à σ , ces estimateurs peuvent s'écrire respectivement $T_n = \frac{\sigma S_n}{E_\sigma(S_n)}$ et $T_n^* = \frac{\sigma S_n^*}{E_\sigma(S_n^*)}$. On sait que T_n est optimal dans la classe des estimateurs sans biais pour le modèle normal considéré, et si on calcule l'efficacité relative asymptotique de T_n^* par rapport à T_n , on trouve qu'elle vaut 0,87.

Cependant, si on considère le modèle contaminé $(1-\epsilon) N(0, \sigma^2) + \epsilon N(0, 9\sigma^2)$, et si on calcule pour ce modèle la limite quand n tend vers l'infini de $\frac{\text{Var}_\sigma(T_n)}{\text{Var}_\sigma(T_n^*)}$, c'est à dire l'efficacité asymptotique relative de T_n^* par rapport à T_n , on trouve une fonction de ϵ qui, au voisinage de 0, est croissante et vaut 1 dès que $\epsilon = 2 \cdot 10^{-3}$.

La performance (asymptotique) de T_n est sensiblement meilleure que celle de T_n^* pour le modèle normal, mais il suffit d'une contamination très faible (indécelable) pour que T_n^* se comporte aussi bien que T_n .

Cet exemple indique simplement la fragilité du caractère optimal face à une petite altération du modèle, mais le nouvel estimateur n'est très bon que pour un type bien particulier de perturbation et ne répond pas, a priori, au problème de la recherche d'estimateurs résistant à une large classe de perturbations. On entend par là des perturbations de natures très diverses, toutes restant cependant petites dans un certain sens ce qui conduit à considérer des voisinages du modèle initial pour des distances convenablement choisies sur l'espace des probabilités.

1b 3. Erreurs d'arrondi :

L'effet des erreurs d'arrondi a fait l'objet de moins d'études que celui des erreurs

grossières. Kempthorne [61] a cependant montré (la justification n'étant malheureusement pas parfaitement rigoureuse), que, si l'on tient compte des erreurs d'arrondi, l'estimateur du maximum de vraisemblance du paramètre θ d'une loi uniforme sur $[0, \theta]$, se converge pas aussi rapidement que ce que l'on prévoit sans en tenir compte ($\frac{1}{n}$ au lieu de $\frac{1}{n^2}$).

2 - VOISINAGES.

Le seul type de modèle étendu que nous ayons vu jusqu'à présent n'est justement pas un voisinage du modèle initial : son seul intérêt dans le cadre de la robustesse est de permettre souvent des calculs explicites et de mettre en lumière l'impact sur les performances des estimateurs d'une infime altération du modèle initial (1.b2 par exemple).

Le modèle initial sera toujours noté $(F_\theta)_{\theta \in \Theta}^{\otimes n}$. Aux paragraphes a, b, c, d, le modèle étendu sera du type $G^{\otimes n}$, $G \in \mathfrak{F}$. Au paragraphe e, des perturbations de l'indépendance sont envisagées.

2a Modèle contaminé.

Si \mathfrak{M} est une classe de lois assez vaste et ϵ un réel de $]0, \frac{1}{2}[$, qui sera généralement assez voisin de 0, on définit le modèle contaminé comme l'ensemble des lois

$$\mathfrak{F}_c(\epsilon) = \{(1-\epsilon)F_\theta + \epsilon H\}_{\theta \in \Theta, H \in \mathfrak{M}}.$$

Ce modèle a été étudié dans Huber [44, 46] et apparaîtra aux Chapitres VII et XIV.

2b Modèle à variation totale.

Si ϵ est un réel positif inférieur à 1, le modèle à variation totale est la réunion, quand θ varie dans Θ , des boules de centre F_θ et rayon ϵ pour la distance en variation totale K . Si P et Q sont deux probabilités sur le même espace (Ω, \mathcal{G}) ; on définit $K(P, Q) = \sup_{A \in \mathcal{G}} |P(A) - Q(A)|$ et par suite le modèle à variation totale est l'ensemble de lois

$$\mathfrak{F}_K(\epsilon) = \{G \mid \exists \theta \in \Theta : K(G, F_\theta) \leq \epsilon\}.$$

Le modèle à variation totale peut être considéré comme une extension du modèle contaminé. En effet, si $G = (1-\epsilon)F + \epsilon H$, $K(F, G) \leq \epsilon$ et, réciproquement, si $K(F, G) = \epsilon$, il existe un réel positif ϵ^* , deux mesures positives H_1 et H_2 (qui ne sont pas, en général, des probabilités) tels que $(1-\epsilon^*)F + \epsilon^* H_1 = (1-\epsilon^*)G + \epsilon^* H_2$. Il suffit de prendre $H_1 = \frac{1}{\epsilon} [F - F \wedge G]$, $H_2 = \frac{1}{\epsilon} [G - F \wedge G]$, $\epsilon^* = \epsilon(1+\epsilon)^{-1}$, et alors $G = F - \epsilon H_1 + \epsilon H_2$.

2c Modèle de Prokhorov.

Soit d la métrique de \mathbf{R} . Si B est un borélien de \mathbf{R} et si ϵ est un réel positif, on note B^ϵ le ϵ voisinage de B , soit $B^\epsilon = \{x \in \mathbf{R} \mid \exists b \in B : d(x, b) < \epsilon\}$.

On définit la distance de Prokhorov π de deux probabilités F et G sur $(\mathbf{R}, \mathfrak{B})$

$$\mathbf{T}(F, G) = \inf \{ \epsilon \mid \forall B \in \mathfrak{B} \quad F(B) \leq G(B^\epsilon) + \epsilon \}.$$

Cette définition semble ne pas être symétrique en F et G , mais l'inégalité qui figure dans le membre de droite implique l'inégalité obtenue en intervertissant les rôles de F et de G , comme on peut le voir en considérant les complémentaires et en remarquant que $B^C \supset ((B^\epsilon)^C)^\epsilon$.

Des deux ϵ qui figurent dans la définition de π , celui qui est relatif à la métrique de \mathbf{R} et qui apparaît dans B^ϵ peut être envisagé comme une diffusion locale de la probabilité due aux erreurs d'arrondi et celui qui est relatif à la probabilité sur $(\mathbf{R}, \mathfrak{B})$ peut représenter la possibilité d'erreurs grossières : une petite fraction ϵ des observations peut se situer n'importe où dans \mathbf{R} .

On choisira l'échelle de mesure de la variable observée de telle sorte que la taille relative de ces deux types d'erreurs soit respectée quand on les exprime par un même nombre positif ϵ .

Le modèle de Prokhorov

$$\mathfrak{F}_\pi(\epsilon) = \{G \mid \exists \theta \in \Theta : \pi(F_\theta, G) \leq \epsilon\}$$

admet pour cas limite le modèle à variation totale : en effet, quand on change d'échelle en prenant une unité de plus en plus petite et tendant vers 0, B^ϵ , pour B fixé, représente un voisinage de B de plus en plus petit et qui tend vers B .

Le modèle à variation totale permet de tenir compte des erreurs grossières mais non des erreurs d'arrondi.

Comme d'autre part la convergence pour la distance de Prokhorov est équivalente à la convergence faible, le modèle \mathfrak{f}_π est celui qui est le plus indiqué puisqu'il permet d'inclure le plus grand nombre de types habituels de déviations.

2d Voisinages pour d'autres distances.

On peut choisir d'autres distances comme par exemple, si on est sur \mathbb{R} muni de la tribu \mathcal{B} de ses boréliens la distance de Lévy λ :

si F et G sont les fonctions de répartition de deux lois sur \mathbb{R} ,

$$\lambda(F, G) = \inf \{ \epsilon \mid \forall t \in \mathbb{R}, G(t) \leq F(t+\epsilon) + \epsilon \}.$$

Ou encore la distance $\delta(F, G) = \sup_{t \in \mathbb{R}} |F(t) - G(t)|$.

2e Perturbations de l'indépendance.

Si l'on envisage maintenant des déviations par rapport à l'hypothèse d'indépendance et d'équidistribution des observations, on est amené à choisir une classe de modèles utilisant la distance de Prokhorov sur \mathbb{R}^n . En fait, si E est un espace métrique complet, muni de sa tribu borélienne \mathcal{B} , si d est la distance, $B \in \mathcal{B}$, $\epsilon > 0$, B^ϵ défini comme en (2c) et P et Q deux probabilités sur (E, \mathcal{B})

$$\pi(P, Q) = \inf \{ \epsilon \mid \forall B \in \mathcal{B} \quad P(B) \leq Q(B^\epsilon) + \epsilon \}.$$

En particulier si $E = \mathbb{R}^n$, on considère le modèle étendu

$$\mathfrak{F}_\pi(\epsilon, n) = \{ Q \mid \exists \theta \in \Theta : \pi(Q, F_\theta^{\otimes n}) \leq \epsilon \}.$$

Bien entendu, cela n'est qu'un type de modèles. D'autres possibilités intéressantes existent, en particulier celles faisant intervenir des types de dépendance particuliers dus au phénomène lui-même (statistique des processus) ou à l'expérience (introduction de dépendances markoviennes par l'expérimentateur).

Catherine HUBER
Département de Mathématique
Université de Paris-Nord
Avenue J.B. Clément
93430 VILLETANEUSE

CHAPITRE III

DÉFINITIONS ET RAPPELS

1- NOTATIONS ET DÉFINITIONS UTILISÉES.

1.a. Lois de probabilité sur \mathbb{R} et fonctions de répartition.

On notera par des lettres $X, Y \dots$ des variables aléatoires réelles (l'espace de probabilité (Ω, \mathcal{A}, P) n'étant pas en général mentionné). La loi de X est l'image de P par X . La fonction de répartition de la variable aléatoire réelle X est la fonction $\mathbb{R} \rightarrow [0, 1]$ définie par $x \mapsto P(X \leq x)$. On notera indifféremment par la même lettre $F, G \dots$, la loi et sa fonction de répartition. F n'étant pas continue en général, on définit sa fonction réciproque F^{-1} par la formule $F^{-1}(t) = \inf \{x, F(x) \geq t\}$, $t \in]0, 1[$; on a alors F^{-1} est croissante, continue à gauche, $(F^{-1}(t) \leq x)$ équivaut à $t \leq F(x)$ pour tout $t \in]0, 1[$, $x \in \mathbb{R}$, et $F^{-1}(F(x)) = x$, F p.s. (en effet, il existe x_t tel que $E_t = \{x, F(x) \geq t\} = [x_t, \infty[$, car si $x \in E_t$, $y \geq x \in E_t$ et si E_t était ouvert, sa borne inférieure x_t serait telle que $F(x_t) < t$ et $F(x) \geq t$ pour tout $x \geq x_t$ ce qui contredit la continuité à droite de F , les assertions ci-dessus sont alors évidentes). Remarquons que si $F^{-1}(F(x)) < x$ alors x est dans un palier de F , si $F(F^{-1}(t)) > t$, t est un saut de F^{-1} .

Parmi les mesures sur \mathbb{R} , nous distinguerons l'ensemble noté \mathcal{M}_n des lois d'échantillonnage, c'est-à-dire l'ensemble des probabilités de la forme

$\frac{1}{n} (\delta_{x_1} + \dots + \delta_{x_n})$, où δ_x est la masse de DIRAC en x , et x_1, \dots, x_n des points de \mathbb{R} , distincts ou non.

Notant dx la mesure de LEBESGUE sur \mathbb{R} , on appelle densité d'une loi absolument continue par rapport à dx , sa dérivée de RADON-NYKODYM. Si F a pour densité f on appelle information de F , la quantité $I(F)$ définie par

$I(F) = \int \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx \leq \infty$ si cette quantité existe, c'est-à-dire si f est partout dérivable, $\frac{f'}{f}(x)$ étant par convention 0 si $f(x) = 0$. Si F n'admet pas une densité dérivable, on pose $I(F) = \infty$.

1.b. Echantillon paramètre de position et d'échelle.

Sauf mention explicite, par n-échantillon d'une loi F , on entend n variables $X_1 \dots X_n$ indépendantes, de loi F . La loi sur \mathbb{R}^n , de vecteur $X = (X_1, \dots, X_n)$ sera notée $F^{\otimes n}$. Si F a une densité f , la fonction de vraisemblance est $f(X_1) \dots f(X_n)$, on travaillera sur son logarithme noté $L(X_1 \dots X_n)$, soit $L(x_1 \dots x_n) = \log f(x_1) + \dots + \log f(x_n)$. Si $\theta \in \mathbb{R}$, $F(x - \theta)$ désigne l'image de F par la translation θ . Si l'on considère une famille du type $\{F(x - \theta), \theta \in \mathbb{R}\}$, θ est appelé paramètre de position, de manière plus générale, pour une famille de type $\{F(\frac{x - \theta}{\sigma}), \theta \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$, θ est appelé paramètre de position et σ paramètre d'échelle.

1.c. Estimateurs.

Suivant l'abus de notation traditionnel en statistique, si $(\hat{\theta}_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires considérées comme des estimateurs de θ à partir d'un n -échantillon, lorsqu'il n'y a pas de confusion possible, on parlera simplement de l'estimateur $\hat{\theta}_n$ et de ses propriétés asymptotiques et l'on écrira $\hat{\theta}_n$ au lieu de $(\hat{\theta}_n)_{n \in \mathbb{N}}$. Soit à estimer le paramètre θ d'une famille $\{F(x, \theta), \theta \in \mathbb{R}\}$. Soit G_n un groupe de transformations sur \mathbb{R}^n , G un groupe de transformations sur \mathbb{R} . $\hat{\theta}_n$ sera dit

invariant (pour le groupe G_n) si pour tout $g \in G_n$, on a

$\hat{\theta}_n(g(X_1 \dots X_n)) = \hat{\theta}(X_1 \dots X_n) \cdot \hat{\theta}_n$ sera dit équivariant (pour le groupe G) si pour tout $g \in G$, $\hat{\theta}_n(g(X_1) \dots g(X_n)) = g(\hat{\theta}(X_1 \dots X_n))$. Les groupes utilisés seront ceux des translations et des homothéties positives sur \mathbb{R}^n .

2 - EFFICACITE ET EFFICACITE RELATIVE ASYMPTOTIQUE (E.R.A.).

2.a. Efficacités des estimateurs.

Soit $\hat{\theta}_n$ un estimateur sans biais équivariant de θ paramètre de translation. On a l'inégalité classique de CRAMER-RAO , $\text{Var } \hat{\theta}_n \geq \frac{1}{n I(F)}$, $I(F)$ information de F . $\hat{\theta}_n$ est dit efficace si $\text{Var } \hat{\theta}_n = \frac{1}{n I(F)}$. Pour les estimateurs équivariants, on a le résultat classique suivant, dû à PITTMAN .

Théorème :

Soit $\{F(x - \theta), \theta \in \mathbb{R}\}$, F donné , alors l'estimateur $\tilde{\theta}_n$ défini par la formule ci-dessous est de variance minimum dans la classe des estimateurs sans biais équivariants par translation. $\tilde{\theta}_n = U - E_0^Y U$, où U est un estimateur quelconque de la classe considérée, et Y est le vecteur aléatoire à $(n - 1)$ dimensions $Y = (X_2 - X_1, \dots, X_n - X_1)$. On peut en particulier choisir $U = \bar{X}$ ou $U = X_1$. (En général l'estimateur de PITTMAN n'est pas efficace).

L'efficacité asymptotique d'un estimateur $\hat{\theta}_n$ est donnée, lorsqu'elle existe par

$$e_\infty = \lim_{n \rightarrow \infty} \frac{e_{\hat{\theta}_n}}{\theta_n} , \quad \text{où} \quad e_{\hat{\theta}_n} = \frac{1}{n I(F) \text{Var } \hat{\theta}_n}$$

Si $e_\infty = 1$, $\hat{\theta}_n$ est dit asymptotiquement efficace.

On peut alors définir, l'efficacité relative asymptotique (E.R.A.) de deux estimateurs $\hat{\theta}_n$, $\hat{\theta}'_n$. En général, on prend comme mesure de l'E.R.A. la limite, lorsqu'elle existe, du rapport des variances des deux estimateurs $\hat{\theta}_n$ et $\hat{\theta}'_n$. En particulier, lorsque $\sqrt{n}(\hat{\theta}_n - \theta)$ et $\sqrt{n}(\hat{\theta}'_n - \theta)$ sont asymptotiquement normaux, l'E.R.A. existe.

L'estimateur de PITTMAN est asymptotiquement d'E.R.A. ≥ 1 par rapport à tout autre estimateur sous des conditions très larges. Il est asymptotiquement efficace sous ces mêmes conditions .

2.b. E.R.A. des tests.

Cette notion est le plus souvent introduite de manière imprécise ; elle est surtout intéressante comme élément indicatif du comportement asymptotique des tests, dans des conditions de régularité convenables.

L'E.R.A. de deux tests de même niveau α , tests unilatères ou bilatères sur \mathbf{R} , peut être alors interprété comme le rapport des écarts $\theta_n - \theta_0$ et $\theta'_n - \theta_0$ entre hypothèse nulle et alternative conduisant à une même puissance β fixée à l'avance.

Soit à tester $\theta - \theta_0$ contre $\theta > \theta_0$, où θ_0 est un paramètre réel et $(F_\theta)_\theta \in \mathbf{R}$ une famille de lois sur \mathbf{R} . On ne s'intéresse qu'à la situation suivante réalisant les hypothèses notées de (*) à (***) .

(*) La région critique de test ϕ est définie à partir d'une statistique T et d'un point critique k par $\phi = 1$ si $T > k$.

$\phi = 0$ si $T < k$, le niveau α étant fixé dans toute la suite, on désigne par γ le nombre $\in (0, 1)$ tel que $\gamma P_{\theta_0}(T = k) + P_{\theta_0}(T > k) = \alpha$.

Dans la suite on s'intéressera à des suites (ϕ_n) , (T_n) de tels tests, n étant la taille de l'échantillon.

(**) On suppose $E_\theta T_n^2 < \infty$ pour tout θ , tout n . On pose $\mu_n(\theta) = E_\theta T_n$, $\sigma_n(\theta) = E_\theta (T_n - \mu_n(\theta))^2$ et on suppose $\lim_n P_{\theta_n} \left(\frac{T_n - \mu_n(\theta)}{\sigma_n(\theta)} < x \right) = \phi(x)$, pour $x \in \mathbf{R}$, où ϕ est la répartition normale.

(***) On pose $\theta_n = \theta_0 + \frac{\tau}{n^\delta}$, pour $\tau > 0$ et $\delta > 0$ fixés. On suppose μ dérivable sur (θ_0, θ_1) , $\theta_1 > \theta_0$, que $\mu'(\theta_0) > 0$ et μ' monotone sur (θ_0, θ_1) . On suppose de plus :

$$\lim_n \frac{\sigma_n(\theta_0)}{\sigma_n(\theta_n)} = 1, \quad \lim_n \frac{\mu'_n(\theta_n)}{\mu'_n(\theta_0)} = 1, \quad \lim_n \frac{\mu'_n(\theta_0)}{n^\delta \sigma_n(\theta_0)} = \ell > 0$$

(***) On suppose de plus que $\lim_n P_{\theta_n} \left(\frac{T_n - \mu(\theta_n)}{\sigma_n(\theta_n)} < x \right) = \phi(x)$ pour tout $x \in \mathbb{R}$.

Ces conditions sont à la fois très restreignantes et pratiquement les seules à être utiles en pratique. Il n'y a pas de conditions générales très agréables pour développer la théorie qui suit.

Etudions la fonction puissance du test T_n au point θ_n , et montrons qu'elle tend vers une limite. Posons $\beta_n(\theta_n) = P_{\theta_n}(T_n > k_n) + \gamma_n P_{\theta_n}(T_n = k_n)$.

Comme $\lim_n P_{\theta_0} \left(\frac{T_n - \mu_n(\theta_0)}{\sigma_n(\theta_0)} < x \right) = \Phi(x)$, si l'on pose $\lambda_n = \frac{k_n - \mu_n(\theta_0)}{\sigma_n(\theta_0)}$, on a $\lim \lambda_n = \lambda_\alpha = \Phi^{-1}(1 - \alpha)$. On peut écrire :

$$\beta_n(\theta_n) = P_{\theta_n} \left(\frac{T_n - \mu_n(\theta_n)}{\sigma_n(\theta_n)} > k_n \right) + \gamma_n P_{\theta_n} \left(\frac{T_n - \mu_n(\theta_n)}{\sigma_n(\theta_n)} = k_n \right)$$

$$\text{Posons } h_n = \lambda_n \frac{\sigma_n(\theta_0)}{\sigma_n(\theta_n)} + \frac{\mu_n(\theta_0) - \mu_n(\theta_n)}{\sigma_n(\theta_n)}$$

Comme T_n est sous P_{θ_n} asymptotiquement gaussienne, l'amplitude de son plus grand saut tend vers zéro et donc $\lim_n \beta_n(\theta_n) = \lim_n P_{\theta_n} \left(\frac{T_n - \mu_n(\theta_n)}{\sigma_n(\theta_n)} > k_n \right)$.

Si $h_n \rightarrow \phi(1 - \beta)$, alors $\beta_n(\theta_n) \rightarrow \beta$ (avec $\beta > \alpha$).

$$\text{Or } h_n = \lambda_n \frac{\sigma_n(\theta_0)}{\sigma_n(\theta_n)} + \frac{\tau}{\delta} \mu'_n(\eta_n \theta_n), \quad 0 < \eta_n < 1.$$

De (*) et (**) on déduit que h_n a même comportement asymptotique que $\lambda_n - \frac{\tau}{n^p} \frac{\mu'_n(\theta_0)}{\sigma_n(\theta_0)}$ qui tend vers $\lambda - \tau \ell$.

On a donc $\lambda - \tau \ell = \phi^{-1}(1 - \beta)$ si $h_n \rightarrow \phi(1 - \beta)$ c'est-à-dire si $\beta_n(\theta_n) \rightarrow \beta$.

On peut alors passer à la notion d'efficacité relative asymptotique (E.R.A.) de 2 tests T_n^1, T_n^2 de même niveau α .

Définition :

Soit un couple fixé, α, β ; $0 < \alpha < \beta < 1$. Soit deux tests ϕ_i , $i = 1, 2$ associés aux statistiques T_i et aux constantes δ_i, ℓ_i telles que les hypothèses (*) à (**) soient vérifiées. On appelle alors E.R.A. de ϕ_1 par rapport à ϕ_2 la quantité

$$\lim_{n \rightarrow \infty} \frac{\tau_1}{n^{\delta_1}} = \frac{n^{\delta_2}}{\tau_2}.$$

L'efficacité peut être comprise entre 0 et $+\infty$.

Si $\delta_1 < \delta_2$, l'efficacité est infinie.

Si $\delta_1 = \delta_2$, elle vaut τ_1/τ_2 (soit ℓ_2/ℓ_1).

Si $\delta_1 > \delta_2$, elle vaut 0.

Remarquons que l'E.R.A. est fonction du couple (α, β) , ce qui est assez gênant.

Cependant il est fréquent d'obtenir une E.R.A. indépendante de ce couple.

CHAPITRE IV

M-ESTIMATEURS

(Maximum de vraisemblance modifié)

par

Claude DENIAU - Georges OPPENHEIM - Claude VIANO

1 - INTRODUCTION.

Etant donné un échantillon (x_1, \dots, x_n) , de taille n , de la loi $F_{\xi_0} = F(x - \xi_0)$, le problème auquel nous sommes confrontés est l'estimation du paramètre ξ_0 de position (la variance étant connue), compliqué par le fait que nous supposons F approximativement connue ; "approximativement connue" signifiant :

- i) la loi F est la ϵ contaminée, $F = (1 - \epsilon)G + \epsilon F$, d'une loi G connue ou
- ii) la loi F appartient à un voisinage \mathcal{M} d'une loi G connue.

S'inspirant des techniques usuelles d'estimation : moindres carrés, maximum de vraisemblance, P.J. HUBER [44] définit une classe d'estimateurs (les M-estimateurs) tous équivariants par translation. Un tel estimateur $T_n(x_1, \dots, x_n)$ est solution de (1)

$$\inf_T \left\{ \sum_{i=1}^n \rho(x_i - T) \right\} \quad (1)$$

où ρ est une fonction convexe telle que $\lim_{|x| \rightarrow \infty} \rho(x) = +\infty$.

Sous certaines conditions et lorsque la loi F est connue l'estimateur du maximum de vraisemblance (solution de (1) avec $\rho = \text{Log } f$, où f est la densité de F) est

asymptotiquement efficace. Cependant dès que F n'est "qu'approximativement connue" cet estimateur est peu robuste en ce sens que $\lim_{n \rightarrow \infty} \text{var} \left\{ \sqrt{n} [T_n - \xi_0] \right\}$ n'est pas bornée en général, dans le domaine \mathcal{M} . On peut penser que la non robustesse des estimateurs usuels provient de la croissance trop rapide de la fonction ρ lorsque $|x|$ tend vers l'infini.

Restreignant notre étude aux lois symétriques, parmi les solutions de (1) nous chercherons des conditions suffisantes sur ρ (ou sur $\psi(x) = \frac{d}{dn} \rho(x)$ lorsqu'elle existe ⁽¹⁾) pour obtenir des estimateurs, convergents, asymptotiquement normaux (C.A.N.).

(1) En tout $x \in \mathbb{R}$, $\psi(x+0)$ et $\psi(x-0)$ existent ; on pose alors $\psi(x) = \frac{\psi(x+0) + \psi(x-0)}{2}$

2 - DÉFINITION DES M-ESTIMATEURS.

Soit ψ une application de \mathbb{R} dans \mathbb{R} monotone non décroissante prenant des valeurs positives et négatives, et F une loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Pour $\xi \in \mathbb{R}$ posons : $\lambda_F(\xi) = \int \psi(x-\xi)F(dx)$; si F_n est la fonction de répartition d'un n -échantillon, alors, $\lambda_{F_n}(\xi) = \frac{1}{n} \sum_{i=1}^n \psi(x_i - \xi)$.

Lemme 1.

S'il existe $\xi_0 \in \mathbb{R}$ tel que $\lambda_F(\xi_0) < \infty$, alors λ_F est définie partout, monotone non croissante, prenant des valeurs positives et négatives.

(Pour une preuve, voir P.J. HUBER [44]).

On peut remarquer que si ψ est bornée, cette condition est réalisée quelle que soit F .

Définition 1.

Soient (x_1, x_2, \dots, x_n) un échantillon de taille n et F_n la fonction de répartition empirique de cet échantillon. Posons :

$$T_n^*(F_n) = \sup \{ \xi \mid \lambda_{F_n}(\xi) > 0 \}$$

$$T_n^{**}(F_n) = \inf \{ \xi \mid \lambda_{F_n}(\xi) < 0 \}$$

Alors

$$T_n^*(F_n) \leq T_n^{**}(F_n).$$

On appelle M-estimateur d'un paramètre de position une application mesurable T_n de \mathcal{F}_n dans \mathbb{R} définie par :

i) si $T_n^*(F_n) = T_n^{**}(F_n)$, on pose $T_n(F_n) = T_n^*(F_n)$,

ii) si $T_n^*(F_n) \neq T_n^{**}(F_n)$, on choisit $T_n(F_n)$ dans $[T_n^*(F_n), T_n^{**}(F_n)]$;

en pratique on pose $T_n(F_n) = \frac{T_n^*(F_n) + T_n^{**}(F_n)}{2}$. Un cas intéressant est celui étudié par P.J. HUBER [45] : $T_n(F_n)$ est randomisé et prend la valeur $T_n^*(F_n)$ avec la probabilité π et $T_n^{**}(F_n)$ avec la probabilité $(1-\pi)$.

Plus généralement on pourra considérer T_n comme la restriction à \mathcal{F}_n d'une fonctionnelle $T : \mathcal{F} \rightarrow \mathbb{R}$ définie en remplaçant λ_{F_n} par λ_F ($F \in \mathcal{F}$) dans le lemme 2.

Propriétés de T_n :

i) L'estimateur T_n peut être considéré comme une application mesurable de \mathbf{R}^n dans \mathbf{R} , invariante par le groupe \mathcal{S}_n des permutations de $\{1, 2, \dots, n\}$

ii) L'estimateur T_n est équivalent par translation i.e.

$$\begin{aligned} x = (x_1, \dots, x_n) \in \mathbf{R}^n ; \underline{a} = (a, \dots, a) \in \mathbf{R}^n \\ T_n(x + \underline{a}) = T_n(x) + a. \end{aligned}$$

Remarque 1.

La propriété ii) permet d'écrire :

Quel que soit $\theta \in \Theta$: $T_n(x + \underline{\theta}) = T_n(x) + \theta$; donc si les x_i ont pour loi F_θ , la loi de $T_n(x + \underline{\theta})$ est indépendante de θ . Nous utiliserons ce résultat dans la démonstration de la normalité asymptotique de T_n en posant $\theta = 0$.

Exemple 1 : Moyenne.

Soit $(x_1, \dots, x_n) \in \mathbf{R}^n$ un échantillon de taille n . La moyenne empirique de l'échantillon : $T_n(x) = \frac{1}{n} \sum_{i=1}^n x_i$ est l'unique nombre réel tel que :

$$\text{Quel que soit } \theta \in \mathbf{R} : \sum_{i=1}^n (x_i - \theta)^2 \geq \sum_{i=1}^n (x_i - T_n(x))^2 \quad (2.1)$$

$$\text{ou : } \sum_{i=1}^n (x_i - T_n(x)) = 0 \quad (2.2)$$

Dans (2.1) intervient la fonction $\rho : x \rightarrow x^2$ (strictement convexe et continûment différentiable) ; dans (2.2) intervient la fonction $\psi = \frac{d\rho}{dx}$.

Exemple 2 : Moyenne Huber k-winzorisée.

Considérons la fonction convexe, continûment différentiable ρ_k suivante :

$$\begin{aligned} \rho_k : \mathbf{R} &\rightarrow \mathbf{R}_+ \\ \rho_k : t &\mapsto \begin{cases} t^2/2 & |t| \leq k \\ k|t| - k^2/2 & |t| > k \end{cases} . \end{aligned}$$

Notons

$$\psi_k = \frac{d}{dt} \rho_k$$

$$\psi_k : t \longrightarrow \begin{cases} t & |t| \leq k \\ k \operatorname{sign} t & |t| > k \end{cases}$$

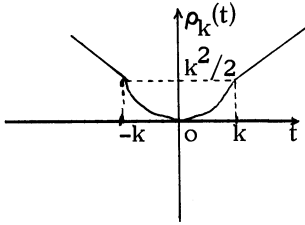


Fig. 1

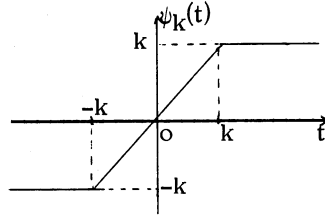


Fig. 2

Les propriétés de ρ_k entraînent l'existence d'un intervalle fermé $[T_n^*(x), T_n^{**}(x)]$ de solutions de l'équation

$$\sum_{i=1}^n \psi_k(x_i - \theta) = 0.$$

Posons :

$$[1] = \{i \in [1, n] ; |x_i - T_n(x)| \leq k\}$$

$$[2] = \{i \in [1, n] ; |x_i - T_n(x)| > k\}.$$

Alors, T_n satisfait à l'équation :

$$\sum_{i \in [1]} (x_i - T_n(x)) + \sum_{i \in [2]} k \operatorname{sign}(x_i - T_n(x)) = 0$$

$$T_n(x) = \frac{1}{\operatorname{Card}[1]} \sum_{i \in [1]} x_i + \frac{k}{\operatorname{Card}[2]} \sum_{i \in [2]} \operatorname{sign}(x_i - T_n(x))$$

Ecrivons $T_n(x)$ de deux manières :

$$i) \quad T_n(x) = 1/n \sum_{i=1}^n y_i \quad \text{avec} \quad y_i = \begin{cases} x & \text{si } i \in [1] \\ T_n(x) + k \operatorname{sign}(x_i - T_n(x)) & \text{si } i \in [2] \end{cases}$$

(les observations sont équipondérées).

$$ii) \quad T_n(x) = \sum_{i=1}^n \alpha_i x_i \quad \text{avec} \quad \alpha_i = \begin{cases} 1 & \text{si } i \in [1] \\ k/|x_i - T_n(x)| & \text{si } i \in [2] \end{cases}$$

(les observations ne sont plus équipondérées).

La détermination de $T_n(x)$ et des valeurs extrêmes ramenées à $\pm k$ est pratiquement obtenue par itérations ; pour plus de détails on peut se reporter à la thèse de D.A. RELLES [78].

On remarquera que :

$$\lim_{k \rightarrow 0} T_n(x) = \text{médiane de l'échantillon}$$

$$\lim_{k \rightarrow \infty} T_n(x) = \text{moyenne de l'échantillon.}$$

On choisira donc des valeurs de k , permettant de prendre en compte les propriétés de robustesse de la médiane et d'efficacité de la moyenne.

3 - CONVERGENCE ET NORMALITÉ ASYMPTOTIQUE DES M-ESTIMATEURS.

Condition suffisante de convergence d'une suite $(T_n)_{n \in \mathbb{N}}$.

Dans tout ce paragraphe $F \in \mathfrak{F}$ est donnée et ψ est une application de \mathbb{R} dans \mathbb{R} , non décroissante prenant des valeurs positives et négatives.

Soit (T_n) une suite de fonctions $T_n : \mathbb{R}^n \rightarrow \mathbb{R}$, mesurables satisfaisant à la définition 1.

Théorème 1.

Supposons que l'application $\xi \mapsto \lambda_F(\xi)$ soit définie (lemme 1) et qu'il existe $\xi_0 \in \mathbb{R}$ tel que :

pour tout $\xi \in]-\infty, \xi_0[: \lambda_F(\xi) > 0$

pour tout $\xi \in]\xi_0, +\infty[: \lambda_F(\xi) < 0$.

Alors (T_n) converge F -presque surement vers ξ_0 .

(La démonstration est immédiate en utilisant la monotonie de ψ , la définition de ξ_0 et la loi forte des grands nombres ; P.J. HUBER [44].

Théorème 2.

i) Supposons que

a) $\xi_0 \in \mathbb{R}$ est tel que $\lambda_F(\xi_0) = 0$.

b) λ_F strictement décroissante et continue dans un voisinage de ξ_0 .

c) $\int \psi^2(x - \xi) F(dx) < \infty$ et continue en $\xi = \xi_0$.

Alors : $\mathcal{L}(\sqrt{n} \lambda_F(T_n - \xi_0)) \xrightarrow{n \rightarrow \infty} \eta(0, \sigma_1^2)$ où $\sigma_1^2 = \int \psi^2(x - \xi_0) F(dx)$.

ii) Supposons λ_F différentiable en ξ_0 et $\lambda'_F(\xi_0) \neq 0$.

Alors : $\mathcal{L}(\sqrt{n}(T_n - \xi_0)) \xrightarrow{n \rightarrow \infty} \eta(0, \sigma_{F, \psi}^2)$ où $\sigma_{F, \psi}^2 = \frac{\sigma_1^2}{[\lambda'_F(\xi_0)]^2}$.

Preuve : Pour alléger les notations posons $\xi_0 = 0$.

i) Il existe une suite (k_n) de nombres réels telle que

$$* \lim_{n \rightarrow \infty} k_n = 0$$

* Pour tout $n > N : \sqrt{n} \lambda_F(k_n) = a$ où a est une constante arbitraire.

En effet, λ_F est bijective dans un voisinage V de zéro, donc il existe $N \in \mathbb{N}$ tel que pour tout $n > N : \frac{a}{\sqrt{n}} \in V$; on pose alors $k_n = \lambda^{-1}(a/\sqrt{n})$.

Démontrons que :

$$\lim_{n \rightarrow \infty} P\{\sqrt{n} \lambda(T_n) > a\} = \Phi\left(-\frac{a}{\sigma_1}\right)$$

où Φ est la fonction de répartition de la loi de Gauss $\mathcal{N}(0,1)$.

Remarquons d'abord que : $\{x \mid \sqrt{n} \lambda(T_n(x)) > a\} = \{x \mid T_n(x) > k_n\}$.

On est ramené à étudier : $\lim_{n \rightarrow \infty} P\{x \mid T_n(x) < k_n\}$.

$$\begin{aligned} \text{Or} \quad \{x \mid T_n(x) < k_n\} &\subset \{x \mid \sum_{i=1}^n \psi(x_i - k_n) \leq 0\} \\ \{x \mid \sum_{i=1}^n \psi(x_i - k_n) < 0\} &\subset \{x \mid T_n(x) \leq k_n\} \end{aligned}$$

et Φ est absolument continue par rapport à la mesure de Lebesgue donc :

$$\begin{aligned} \lim_{n \rightarrow \infty} P\{x \mid T_n(x) < k\} &= \lim_{n \rightarrow \infty} P\{x \mid \sum_{i=1}^n \psi(x_i - k_n) < 0\} \\ &= \lim_{n \rightarrow \infty} P\{x \mid \frac{1}{\sqrt{n}} \sum_{i=1}^n [\psi(x_i - k_n) - \lambda_F(k_n)] < -a\} \\ &= \lim_{n \rightarrow \infty} P\{x \mid \frac{1}{\sqrt{n}} \sum_{i=1}^n y_{n,i} < -a\} \end{aligned}$$

en posant $y_{n,i} = \psi(x_i - k_n) - \lambda_F(k_n)$. La première partie sera démontrée si l'on peut appliquer un théorème central limite à la somme de variables aléatoires indépendantes équadistribuées $\sum_{i=1}^n y_{n,i}$.

Nous sommes ramenés à vérifier des conditions de Lindeberg (voir par exemple LOEVE [70] p. 295).

Soit à montrer que :

$$\forall \epsilon > 0 : \lim_{n \rightarrow \infty} \frac{1}{\sigma_n^2} \sum_{i=1}^n \int_{E_n} y_{n,i}^2 F(dx) = 0 \quad \text{où} \quad \begin{cases} E_n = \{x \mid |y_{n,i}| > \epsilon \sqrt{n}\} \\ \sigma_n^2 = \sum_{i=1}^n \int y_{n,i}^2 F(dx) = n \sum_{i=1}^n y_{i,n}^2 \end{cases}$$

ou ce qui est équivalent : $\forall \epsilon > 0 \quad \lim_{n \rightarrow \infty} \int_{E_n} y_{n,i}^2 F(dx) = 0 \quad (3.1)$

Par le théorème de la convergence dominée et la continuité de λ_F au voisinage de $\xi = 0$, on déduit que (3.1) est vérifiée si :

$$\lim_{n \rightarrow \infty} \int_{E_n'} \psi^2(x_i - k_n) F(dx) = 0 \quad \text{où} \quad E_n' = \{x \mid |\psi(x_i - k_n)| > \epsilon \sqrt{n}\} \quad (3.2)$$

Enfin : $\forall \delta > 0, \exists n_0 \in \mathbb{N}, \forall n \geq n_0 : |k_n| < \delta$. Posons

$$u(x_i) = \text{Max} \{ |\psi(x_i - \delta)|, |\psi(x_i + \delta)| \}.$$

Alors pour tout $n \geq n_0 : |\psi(x_i - k_n)| \leq u(x_i)$. Posons

$$E_n'' = \{x \mid u(x_i) > \sqrt{n} \epsilon\} ; \text{ et pour } n \geq n_0, E_n' \subset E_n''.$$

Donc pour $n \geq n_0 : \int_{E_n'} \psi^2(x_i - k_n) F(dx) \leq \int_{E_n''} u^2(x_i) F(dx)$. Et de la continuité en

$\xi = 0$ de $\int \psi^2(x - \xi) F(dx)$ on déduit :

$$\lim_{n \rightarrow \infty} \int_{E_n''} u^2(x_i) F(dx) = 0$$

ce qui achève la démonstration.

ii) L'application λ_F étant différentiable en $\xi_0 = 0$, on peut écrire :

$$\begin{aligned} \lambda_F(T_n) &= T_n \lambda_F'(0) + |T_n| o(T_n) \\ \frac{\sqrt{n} \lambda_F(T_n)}{\sqrt{n} T_n \lambda_F'(0)} &= 1 + \frac{o(T_n)}{\lambda_F'(0)} \end{aligned} \quad (3.3)$$

Or (T_n) converge F-presque sûrement vers 0, donc le second membre de (3.3) converge vers 1 en probabilité

donc :

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathcal{L}(\sqrt{n} \lambda_F(T_n)) &= \eta(0, \sigma_1^2) \\ \lim_{n \rightarrow \infty} \mathcal{L}(\sqrt{n} T_n) &= \eta(0, \sigma_{F, \psi}^2) \quad \text{où} \quad \sigma_{F, \psi}^2 = \frac{\sigma_1^2}{[\lambda'(0)]^2}. \end{aligned}$$

(Pour ce résultat classique, voir par exemple H. CRAMER [21] p. 254).

Remarque 2.

i) Si ψ est suffisamment régulière (voir les Huber k -winzorisés, où des exemples dans ANDREWS et AL [2]) : $\lambda_F'(0)$ existe quelle que soit F .

ii) Si F est telle que son information de Fisher $I(F)$ soit finie, alors quelle que soit ψ monotone : $\lambda_F^1(0) = \int \psi(x) f'(x) dx$ (où f est la densité de F par rapport à la mesure de Lebesgue) et $\sigma_{F,\psi}^2 \geq \frac{1}{I(F)}$, avec égalité si et seulement si $\psi_V = k \frac{f'}{f}$ (on ne demande plus la monotonie de ψ_V). La suite (T_n) de M-estimateurs associés à ψ_V est une suite d'estimateurs du maximum de vraisemblance et ces estimateurs sont asymptotiquement efficaces.

Claude DENIAU
Georges OPPENHEIM
Claude VIANO

U.E.R. Mathématiques
Université Paris V
12, Rue Cujas
75005 PARIS

CHAPITRE V

ESTIMATION D'UN PARAMÈTRE DE TRANSLATION A PARTIR DE TESTS DE RANG

par

Robert AZENCOTT.

1 - MODÈLE PROBABILISTE.

On observe n variables aléatoires $X_1 \dots X_n$ définies sur un espace de probabilité $(\Omega, \mathcal{A}, P_\theta)$ où P_θ est une probabilité sur Ω dépendant du paramètre réel θ que l'on cherche à estimer. On suppose que sous P_θ , les X_i sont indépendantes et de même loi F_θ définie par

$$F_\theta(x) = P_\theta(X_i \leq x) = F(x - \theta) \quad n \in \mathbb{R}$$

où F est la fonction de répartition d'une probabilité fixe sur \mathbb{R} .

On suppose que la loi F est symétrique par rapport à 0, c'est-à-dire $F(x) + F(-x) = 1$, et fixée - mais pas nécessairement connue a priori - dans le plus grande partie de cet exposé. Les résultats obtenus ainsi seront ensuite appliqués au cas où F se trouve dans un certain "voisinage" de F_0 , où F_0 est connue.

Les estimateurs de θ que nous présentons ici sont construits à partir de tests de rang et ont été proposés par HODGES et LEHMANN [41], puis étudiés par HODGES [38], BICKEL et HODGES [7], et JAECKEL [58].

2 - CONSTRUCTION D'ESTIMATEURS À MÉDIANE NON BIAISÉE.

Soit $h : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction mesurable servant à tester $\{\theta = 0\}$ contre $\{\theta > 0\}$ de la façon suivante : on rejette $\theta = 0$ si $h(X_1 \dots X_n) \geq C$.

Pour simplifier l'écriture nous noterons

$$h(X) = h(X_1, \dots, X_n) \quad h(X+a) = h(X_1 + a, \dots, X_n + a)$$

Introduisons l'hypothèse.

- (2.1) - $h(X+a)$ est croissante au sens large en a , pour tout X fixé.
 - Sous $\theta = 0$, $h(X)$ a une loi symétrique par rapport à zéro, quelle que soit F .

Posons $\epsilon_n = P_0 \{h(X) = 0\}$.

Pour $a \in \mathbb{R}$, considérons le test T_a de $\{\theta = a\}$ contre $\{\theta > a\}$ qui rejette $\{\theta = a\}$ si $h(X-a) \geq 0$. Les T_a sont de même niveau $\alpha = \frac{1}{2} + \frac{\epsilon_n}{2} = P_a [h(X-a) \geq 0]$.
 Considérons les intervalles de confiance $S(X)$ associés, cf. [68], définis par $S(X) = \{a \mid T_a \text{ accepte } \theta = a \text{ si } X \text{ est observé}\}$.

On a alors $P_\theta \{\theta \in S(X)\} \geq 1 - \alpha = \frac{1}{2} - \frac{\epsilon_n}{2}$

Posons

$$(2.2) \quad a^* = \inf \{a \mid h(X-a) < 0\}$$

On obtient $S(X) \subset [a^*, +\infty[$ et donc

$$P_\theta (a^* \leq \theta) \geq \frac{1}{2} - \frac{\epsilon_n}{2}$$

En considérant les tests T'_a de $\{\theta = a\}$ contre $\{\theta < a\}$ qui rejettent $\{\theta = a\}$ si $h(X-a) \leq 0$ on est de même amené à poser

$$(2.3) \quad a^{**} = \sup \{a \mid h(X-a) > 0\}$$

ce qui donne

$$P_{\theta} (a^{**} \geq \theta) \geq \frac{1}{2} - \frac{\epsilon_n}{2}$$

On a $a^{**} \leq a^*$ et l'estimateur

$$(2.4) \quad \hat{\theta} = \frac{1}{2} (a^{**} + a^*)$$

vérifie

$$(2.5) \quad P_{\theta} (\hat{\theta} \geq \theta) \geq \frac{1}{2} - \frac{\epsilon_n}{2} \quad P_{\theta} (\hat{\theta} \leq \theta) \geq \frac{1}{2} - \frac{\epsilon_n}{2}$$

A tout test h de $\{\theta = 0\}$ contre $\{\theta > 0\}$ vérifiant (2.1) nous associons l'estimateur $\hat{\theta}$ de θ défini par (2.2), (2.3), (2.4). Dans les exemples ci-dessous on a $\lim_{n \rightarrow +\infty} \epsilon_n = 0$ de sorte que $\hat{\theta}$ est "presque" la médiane de $\hat{\theta}$ pour n grand. Sous hypothèses supplémentaires faibles (la normalité asymptotique de $\hat{\theta}$ suffit amplement !) la médiane de $\hat{\theta}$ converge en probabilité vers θ .

3 - TRANSFERT DES PROPRIÉTÉS D'OPTIMALITÉ DE h A $\hat{\theta}$.

3.1. Lemme : (cf. [41]).

Si h vérifie (2.1), si F est continue, les fonctions de répartition de a^* et a^{**} sous P_θ sont continues.

Preuve :

Posons $W_1 = X_1$, $W_i = X_i - X_1$ $i \geq 2$. La fonction $f(z) = h(z, z + W_2, \dots, z + W_n)$ est croissante en z . Soit $\lambda(W_2 \dots W_n) = \sup \{z \mid f(z) < 0\}$, ce qui entraîne $a^* = W_1 - \lambda(W_2 \dots W_n)$. Pour $c \in \mathbb{R}$ on a donc $P_\theta(a^* = c) = P_\theta\{W_1 = \lambda(W_2 \dots W_n) + c\}$.

La loi η de X_1 sous P_θ ne charge pas les points, d'où

$$P_\theta\{W_1 = \lambda(W_2 \dots W_n) + c \mid W_2, \dots, W_n\} = \eta[\lambda(W_2 \dots W_n) + c] = 0$$

ce qui implique $P_\theta(a^* = c) = 0$. Même argument pour a^{**} .

3.2. Lemme (cf. [41]).

Si h vérifie (2.1), si F est continue on a pour tout $u \in \mathbb{R}$, $\theta \in \mathbb{R}$

$$P_\theta\{h(X - u) < 0\} \leq P_\theta\{\hat{\theta} < u\} \leq P_\theta\{h(X - u) \leq 0\}$$

Preuve :

D'après (2.4), (2.3), (2.2) on a

$$\{\hat{\theta} < u\} \subset \{a^{**} < u\} \subset \{h(X - u) \leq 0\}$$

$$\{h(X - u) < 0\} \subset \{a^* \leq u\}$$

D'après le lemme 3.1, on a, à des ensembles P_θ -négligeables près

$$\{a^* \leq u\} = \{a^* < u\} \subset \{\hat{\theta} < u\}$$

3.3. Lemme (cf. [41]).

Soit $h_n = \mathbb{R}^n \rightarrow \mathbb{R}$ une suite de tests vérifiant (2.1). Supposons F continue. Supposons que pour tout $u \in \mathbb{R}$ il existe des suites $\mu_n(u) \in \mathbb{R}$ et $\sigma_n > 0$, avec $\mu_n(0) = 0$ telles que la loi de $\frac{1}{\sigma_n} [h_n - \mu_n(u)]$ sous $P_{u/\sqrt{n}}$ tende faiblement vers $N(0, 1)$.

Supposons que $\lim_{n \rightarrow +\infty} \left(+ \frac{\mu_n(u)}{\sigma_n} \right) = uA$ pour tout $u \in \mathbb{R}$, avec $A \neq 0$. Les $\hat{\theta}_n$ associés aux h_n sont alors asymptotiquement normaux et

$$(3.4) \quad \lim_{n \rightarrow +\infty} P_{\theta} \{ \sqrt{n} (\hat{\theta}_n - \theta) < u \} = \Phi(+uA), \quad u \in \mathbb{R}$$

Preuve :

Par construction, on a pour tout borélien $B \subset \mathbb{R}$,
 $P_{\theta} \{ (\hat{\theta}_n - \theta) \in B \} = P_0 \{ \hat{\theta}_n \in B \}$. On peut donc supposer $\theta = 0$. La définition des P_{θ} donne

$$\begin{aligned} P_0 \{ h_n(X - \frac{u}{\sqrt{n}}) \leq 0 \} &= P_{-\frac{u}{\sqrt{n}}} \{ h_n(X) \leq 0 \} \\ &= P_{-\frac{u}{\sqrt{n}}} \left\{ \frac{1}{\sigma_n} [h_n(X) - \mu_n(-u)] \leq -\frac{\mu_n(-u)}{\sigma_n} \right\} \end{aligned}$$

Les hypothèses sur h_n entraînent donc

$$\lim_{n \rightarrow +\infty} P_0 \{ h_n(X - \frac{u}{\sqrt{n}}) \leq 0 \} = \phi(+uA)$$

et ce résultat subsiste si on remplace ≤ 0 par < 0 . En tenant compte de l'encadrement de $P_0(\sqrt{n} \hat{\theta}_n < u)$ fourni par le lemme 3.2, on obtient (3.4).

Q.E.D.

3.5. Proposition (cf. [41]).

Considérons deux suites (h_n) , (h'_n) de tests vérifiant les hypothèses du lemme 3.3 ; supposons que les constantes A, A' associées soient de même signe. L'efficacité asymptotique relative ("A.R.E." cf. Chap. n°3) de (h_n) par rapport à (h'_n) est

égale à l'efficacité relative des suites d'estimateurs associés $(\hat{\theta}_n)$ et $(\hat{\theta}'_n)$.

Preuve :

Remarquons que dans les exemples $\mu_n(u)$ est l'espérance de h_n sous $P_{u/\sqrt{n}}$, et donc d'après (2.1) on a $A, A' \geq 0$.

Soient μ'_n, σ'_n, A' les quantités associées à (h'_n) par les hypothèses de 3.3. Les variances asymptotiques de $\sqrt{n} \hat{\theta}_n$ et $\sqrt{n} \hat{\theta}'_n$ sont resp. $\frac{1}{A^2}$ et $\frac{1}{A'^2}$ d'après (3.4) ; l'efficacité relative de $(\hat{\theta}_n)$ par rapport à $(\hat{\theta}'_n)$ est donc A^2/A'^2 .

Soit $k_{1-\alpha}$ le $(1-\alpha)$ -quantile de $N(0, 1)$ et soit $C_n = \sigma_n k_{1-\alpha}$. Le test T_n de $\{\theta = 0\}$ contre $\{\theta = \frac{u}{\sqrt{n}}\}$ qui rejette $\{\theta = 0\}$ si $h_n(X) \geq C_n$ est de niveau $\alpha_n = P_0 \{h_n(X) \geq C_n\}$ d'où $\lim_{n \rightarrow +\infty} \alpha_n = \alpha$. La puissance limite de T_n est

$$\beta = \lim_{n \rightarrow +\infty} \beta_n = \lim_{n \rightarrow +\infty} P_{\frac{u}{\sqrt{n}}} \{h_n(X) \geq C_n\} = \Phi(Au + k_{1-\alpha})$$

Le test $T'_{n'}$ de $\theta = 0$ contre $\theta = \frac{u'}{\sqrt{n'}}$ qui rejette $\theta = 0$ si $h'_{n'}(X) \geq \sigma'_{n'} k_{1-\alpha}$ est de niveau asymptotique α et de puissance asymptotique $\phi(A'u' + k_{1-\alpha})$. Clairement, T_n et $T'_{n'}$ auront même puissance asymptotique et même niveau asymptotique contre les mêmes alternatives si $\frac{u}{\sqrt{n}} = \frac{u'}{\sqrt{n'}}$ et $Au = A'u'$, donc si $\frac{n'}{n} = \frac{A^2}{A'^2}$ ce qui par définition indique que l'efficacité asymptotique relative de (h_n) par rapport à $(h'_{n'})$ est $\frac{A^2}{A'^2}$.

Q.E.D.

Les "bons" tests (h_n) fourniront donc de "bons" estimateurs $\hat{\theta}_n$.

4 - LES TESTS DE RANG :

4.1. Un test $h(X)$ de $\{\theta = 0\}$ contre $\{\theta \neq 0\}$ est un test de rang si (cf. [31]) on peut écrire $h(X_1 \dots X_n) = H(R_1, \dots, R_n, \text{sign } X_1 \dots \text{sign } X_n)$ où R_i est le rang de $|X_i|$ dans l'échantillon ordonné associé à $|X_1| \dots |X_n|$, et où $\text{sign } X_i$ vaut 1, 0, -1 resp. lorsque $X_i > 0$, $X_i = 0$, $X_i < 0$.

Les vertus de ces tests sont lucidement décrites dans [31] dont nous extrayons les deux résultats suivants :

4.2. Théorème [31] :

Supposons que F ait une densité f (par rapport à la mesure de LEBESGUE) que f soit absolument continue, et que $\int |f'(x)| dx < +\infty$. Pour F donnée et n fixé, il existe un test de rang h_n localement plus puissant que tous les tests de rang ;

h_n est de type

$$h_n(X) = \sum_{i=1}^n a_n(R_i) \text{sign } X_i$$

avec $a_n(i) = E \left[-\frac{f'}{f} \circ F^{-1} \left(\frac{1}{2} + \frac{1}{2} U^{(i)} \right) \right]$ où $U^{(1)} \dots U^{(n)}$ sont les statistiques d'ordre d'un échantillon de taille n extrait d'une population de loi uniforme sur $[0, 1]$.

Remarquons que si $\frac{f'}{f} \circ F^{-1}$ est dans $L^2[0, 1]$ c'est-à-dire si l'information de FISHER $I(F) = \int \left(\frac{f'}{f}\right)^2 dF$ est finie, le test h_n ci-dessus est asymptotiquement équivalent à h'_n défini comme h_n à partir des scores $a'_n(i) = -\frac{f'}{f} \circ F^{-1} \left(\frac{1}{2} + \frac{1}{2} \frac{i}{n} \right)$ (cf. [31]).

4.3. Théorème [31] :

Supposons $I(F) < +\infty$. Soit θ_n une suite d'alternatives telles que $|\theta_n| \leq \frac{Cte}{\sqrt{n}}$. Pour tout $\alpha \in [0, 1]$, il existe une suite h_n de tests de rang de $\{\theta = 0\}$ contre $\{\theta = \theta_n\}$, de niveau asymptotique α , qui sont asymptotiquement plus puissants que n'importe quelle suite de tests (h'_n) de même niveau asymptotique (les h'_n ne sont pas restreints à être des tests de rang).

Les théorèmes 4.2 et 4.3 montrent l'intérêt des tests de rang du type

$$(4.4) \quad h_n(X_1 \dots X_n) = \sum_{i=1}^n b_n(R_i) \text{ sign } X_i$$

Remarquons que sous P_0 un tel test est de loi symétrique par rapport à 0, quel que soit F ; en effet, la loi de X sous P_0 est invariante par la symétrie autour de 0 dans \mathbb{R}^n , qui transforme $h_n(X)$ en $h_n(-X) = -h_n(X)$. Par contre $h_n(X+u)$ n'est pas croissante en u si les scores $b_n(i)$ sont quelconques ; une vérification longue, mais élémentaire, montre que, pour n fixé, $h_n(X+u)$ est croissante en u si la fonction $i \rightarrow b_n(i)$ est croissante et positive (au sens large).

Dans ces conditions h_n vérifie (2.1) et l'estimateur $\hat{\theta}_n$ de θ associé à h_n est appelé un R-estimateur de θ .

Il existe une autre présentation des tests h_n de type 4.4, par analogie avec le "problème de deux échantillons" (cf. [58]) :

4.5. Lemme :

Soit S_i le rang de X_i dans l'échantillon de taille $2n$ obtenu en ordonnant $\{X_1 \dots X_n, -X_1, \dots, -X_n\}$. Etant donnés des nombres $b(i)$, $i = 1 \dots n$, définissons $J(k)$, $k = 1 \dots 2n$ par

$$-J(n+1-i) = J(n+i) = b(i) \quad i = 1 \dots n$$

Si $F \ll \text{LEBESGUE}$, on a $\sum_{i=1}^n b(R_i) \text{ sign } X_i = \sum_{i=1}^n J(S_i)$, P_θ -p.s. quel que soit θ .

Preuve :

Soit $|X_i(k)|$ le k -ème terme de l'échantillon ordonné associé à $|X_1| \dots |X_n|$; les $X_1 \dots X_n - X_1 \dots - X_n$ sont distincts avec P_θ -probabilité 1. Si $X_i(k) > 0$ on a $S_i(k) = R_i(k) + n$; si $X_i(k) < 0$, on a $S_i(k) = n + 1 - R_i(k)$. Comme $X_j \neq 0$ P_θ -p.s., on obtient :

$$J(S_i(k)) = b(R_i(k)) \text{ sign } X_i(k) \quad P_\theta - \text{p.s.}$$

d'où le résultat puisque $k \rightarrow i(k)$ est une permutation de $1 \dots n$.

Q.E.D.

5 - CALCUL DES R-ESTIMATEURS - EXEMPLES.

5.1. Lemme (cf. [58]).

Etant donnés des scores $b(i)$, $i = 1 \dots n$ posons $d_i = b(i) - b(i-1)$, $i = 2 \dots n$ et $d_1 = b(1)$. Soit $Y_1 \dots, Y_n$ l'échantillon ordonné associé aux X_i . Si $F \ll \text{LEBESGUE}$, on a P_θ - p.s.

$$\sum_{X_i > 0} b(R_i) = \sum_{j=1}^n \sum_{1 \leq k \leq j} d_{1+j-k} I_{\{Y_j + Y_k > 0\}}$$

Preuve :

Si $Y_j < 0$, on a $Y_j + Y_k < 0$ pour tout $k \leq j$ et donc $\sum_{1 \leq k \leq j} d_{1+j-k} I_{\{Y_j + Y_k > 0\}} = 0$. Si $Y_j > 0$ et $k \leq j$, les conditions $\{Y_j + Y_k > 0\}$ et $\{|Y_j| > |Y_k|\}$ sont équivalentes P_θ - p.s. et donc équivalent à $1 + j - k \leq r$ où r est le rang de $|Y_j|$ dans l'échantillon ordonné associé à $|X_1| \dots |X_n|$. Si on note $Y_j = X_{i(j)} > 0$ on a $r = R_{i(j)}$ et donc

$$\sum_{1 \leq k \leq j} d_{1+j-k} I_{\{Y_j + Y_k > 0\}} = \sum_{1 \leq 1+j-k \leq R_{i(j)}} d_{1+j-k} = b[R_{i(j)}]$$

On obtient bien le résultat annoncé car $j \rightarrow i(j)$ est une permutation de $1 \dots n$.

5.2. Proposition (cf. [58]).

Supposons les scores $b(i) \geq 0$, croissants au sens large et $F \ll \text{LEBESGUE}$. Les d_i sont comme en 5.1, et on note $c = \sum_{i=1}^n b(i)$. Soit $\hat{\theta}$ le R-estimateur défini par $h(X) = \sum_{i=1}^n b(R_i) \text{sign } X_i$. Soit Y_i , $1 \leq i \leq n$ les statistiques d'ordre de $(X_1 \dots X_n)$. Pour chaque n-tuple $X_1 \dots X_n$ définissons une probabilité $\nu_{X_1 \dots X_n}$ sur R par

$$\nu_{X_1 \dots X_n} = \frac{1}{c} \sum_{j=1}^n \sum_{1 \leq k \leq j} d_{1+j-k} \frac{\delta_1}{2}(Y_j + Y_k)$$

où δ_x = masse unité en x . Alors $\hat{\theta}(X_1 \dots X_n)$ est la médiane de $\nu_{X_1 \dots X_n}$.

Preuve :

Posons $h' = \sum_{X_i > 0} b(R_i)$, de sorte que $h = 2h' - c$. Le lemme 5.1 entraîne
 $h'(X - a) = c \nu_{X_1 \dots X_n}([a, +\infty[)$. D'après (2.2) on a alors

$$a^* = \inf \{a \mid h'(X - a) < \frac{c}{2}\} = \inf \{a \mid \nu_{X_1 \dots X_n}([a, +\infty[) < \frac{1}{2}\}$$

et de même $a^{**} = \sup \{a \mid \nu_{X_1 \dots X_n}([a, +\infty[) > \frac{1}{2}\}$. Donc $\hat{\theta} = \frac{1}{2}(a^{**} + a^*)$
 est médiane de $\nu_{X_1 \dots X_n}$.

Q.E.D.

5.3. Estimateur de HODGES-LEHMANN (cf. [41], [38]) :

Considérons le test de rang associé par 4.4 aux scores $b(i) = i$ (et donc aux scores $J(n+i) = -J(n+1-i) = i$ dans la présentation 4.5). Alors
 $h = 2 \left(\sum_{X_i > 0} R_i \right) - \frac{n(n+1)}{2}$ est l'analogue du test de WILCOXON ; ce test est optimal parmi les tests de rang, et donc asymptotiquement optimal si $F(x) = (1 + e^{-x})^{-1}$ est la distribution logistique (cf. [31] et 3.5, 4.2, 4.3). L'estimateur $\hat{\theta}$ associé est la médiane de l'ensemble des nombres $M_{jk} = \frac{1}{2}(Y_j + Y_k)$ où $1 \leq k \leq j \leq n$ (notations de 5.2) ; nous le noterons H-L.

5.4. Estimateur de BICKEL-HODGES [7], [38] :

C'est une version simplifiée de H-L, définie comme la médiane des $M_{i, n+1-i}$, $i = 1 \dots n$ (notations de 5.3). Il ne peut pas s'obtenir à partir d'un test de rang de type 4.4 par la méthode ci-dessus, mais c'est bien l'estimateur à "médiane non biaisée" associé au test de GALTON [37] $h = \text{cardinal } \{i \mid M_{i, n+1-i} > 0\}$.

5.5. Estimateur "de VAN DER WAERDEN" :

Considérons le test de rang optimal lorsque F est normal, donné par (cf. [31] et

3.5, 4.2, 4.3) les scores $b_n(i) = E[|V|^{(i)}]$ où $|V|^{(i)}$ est le i -ème statistique d'ordre de l'échantillon $|V_1| \dots |V_n|$, les V_i étant indépendants de loi $N(0, 1)$. Pour n grand, les $b(i)$ peuvent sans perte de puissance asymptotique (cf. 4.2) être remplacés par $\tilde{b}_n(i) = \Phi^{-1}(\frac{1}{2} + \frac{1}{2} \frac{i}{n})$. Les estimateurs $\hat{\theta}_n$ associés n'ont pas de forme plus simple que celle donnée en 5.2, et, on le verra plus bas, sont efficaces quand $F = N(0, 1)$.

5.6. Médiane :

Considérons le test "de signes", optimal lorsque $\frac{dF}{dx}(x) = \frac{1}{2} \exp(-|x|)$, correspondant aux scores $b(i) \equiv 1$; on a $h = \text{card}\{i \mid X_i > 0\}$ et l'estimateur associé $\hat{\theta}$ est la médiane des X_i $1 \leq i \leq n$.

6 - NORMALITÉ ASYMPTOTIQUE.

On a vu que la normalité asymptotique de $\hat{\theta}_n$ sous P_θ peut se déduire de celle de h_n sous P_u / \sqrt{n} . Celle-ci est dans de nombreux cas conséquence d'un théorème profond de LE CAM-HAJEK-SIDAK basé sur les lemmes classiques de LE CAM concernant les hypothèses contigües. Pour la démonstration de 6.1 nous préférons renvoyer à [64] [31].

6.1. Théorème [64] [31] :

Supposons que $I(F) < +\infty$. Soit (φ_n) une suite de fonctions en escalier sur $[0,1]$, avec φ_n constante sur $[\frac{i}{n}, \frac{i+1}{n}[$. Supposons que $\varphi_n \rightarrow \varphi$ dans $L_2[0,1]$. Soit h_n le test de type 4.4 associé aux scores $b_n(i) = \varphi_n(i/n)$. Alors sous P_u / \sqrt{n} , h_n est asymptotiquement normal.

Posons $\tilde{f}(t) = \frac{f'}{f} \circ F^{-1}(\frac{1}{2} + \frac{1}{2}t)$ $t \in [0,1]$ où $f = \frac{dF}{dx}$ et notons $\|\cdot\|_2, \langle \cdot, \cdot \rangle$ la norme et le produit scalaire de $L_2[0,1]$. On a pour tout $a \in \mathbb{R}, u \in \mathbb{R}$

$$\lim_{n \rightarrow +\infty} P_{u/\sqrt{n}} \left\{ \frac{1}{\sigma_n} [h_n - \mu_n(u)] \leq a \right\} = \Phi(a)$$

avec

$$(6.2) \quad \sigma_n = \sqrt{n} \|\varphi\|_2 \quad \mu_n(u) = u \sqrt{n} \langle \varphi, \tilde{f} \rangle$$

6.3. Corollaire (hypothèses, notations, de 6.1) :

Supposons de plus φ_n croissante au sens large pour chaque n . Les R-estimateurs $\hat{\theta}_n$ associés aux tests h_n vérifient $\lim_{n \rightarrow +\infty} P_\theta \{ \sqrt{n} (\hat{\theta}_n - \theta) \leq u \} = \Phi(\frac{u}{\sigma})$ pour tout $u \in \mathbb{R}$ avec $\sigma = \frac{\|\varphi\|_2}{\langle \varphi, \tilde{f} \rangle}$, pourvu que $\langle \varphi, \tilde{f} \rangle \neq 0$.

Preuve :

Appliquer 3.4 et 6.1.

6.4. Remarque :

De 6.2 on tire $\sigma \geq \frac{1}{\|\tilde{f}\|_2} = \frac{1}{I(F)^{1/2}}$ et on a $\sigma = I(F)^{-1/2}$ si et seulement si $\varphi = \tilde{f}$. Comme $\tilde{f} \in L_2[0, 1]$, les scores $\tilde{b}_n(i) = \frac{1}{n} \int_{i/n}^{(i+1)/n} \tilde{f}(t) dt$ vérifient les hypothèses de 6.1, avec $\varphi = \tilde{f}$. Mais en général, ils ne croissent pas avec i et ne sont pas positifs. Cependant si $\text{Log } f$ est concave et atteint son maximum au point 0, les $\tilde{b}_n(i)$ vérifient les hypothèses de 6.2, et les estimateurs $\hat{\theta}_n$ associés sont (d'après l'inégalité de CRAMER-RAO) de variance asymptotique minimale.

6.5. Remarque :

Sous les hypothèses des théorèmes 6.1, 6.2 considérons les scores $\gamma_n(k)$, $k = 1 \dots 2n$ associés aux $b_n(i)$ par $\gamma_n(n+i) = -\gamma_n(n+1-i) = b_n(i)$, c'est-à-dire relatifs à l'échantillon symétrisé (cf. 4.5). On a $\gamma_n(k) = J_n\left(\frac{k}{2n+1}\right)$, où J_n est une fonction en escalier sur $[0, 1]$, constante sur $\left[\frac{k}{2n+1}, \frac{k+1}{2n+1}\right]$ et où dans $L_2[0, 1]$, $\lim_{n \rightarrow \infty} J_n = J$ avec

$$-J\left(\frac{1}{2} - \frac{1}{2}t\right) = J\left(\frac{1}{2} + \frac{1}{2}t\right) = \varphi(t) \quad \text{pour } t \in]0, 1[$$

(pour $t = 0$ on pose $J(0) = 0$). La variance asymptotique σ des estimateurs $\hat{\theta}_n$ associés aux scores "symétrisés" $J_n\left(\frac{i}{2n+1}\right)$, est donnée par 6.2 et on a aisément

$$\sigma = \frac{\|J\|_2}{\langle J, g \rangle} \quad \text{où } \tilde{g}(u) = -\frac{f'}{f} \circ F^{-1}(u). \quad \text{Le } J \text{ optimal (si } \text{Log } f \text{ est concave) est } J = \tilde{g}.$$

Si $\varphi \in L_2[0, 1]$ est de plus croissante, dérivable sur $]0, 1[$ et vérifie $\varphi(0) = 0$, on vérifie que

$$\langle \varphi, \tilde{f} \rangle = \langle J, \tilde{g} \rangle = \int J' \circ F(x) f(x)^2 dx$$

(intégrer par parties si φ est de support compact, approcher φ par η_k où

$\lim_{k \rightarrow \infty} \eta_k = 1$, η_k est à support compact $\|\eta_k'\|_\infty \leq \text{cte}$). On obtient donc pour la variance asymptotique de $\hat{\theta}_n$ la formule (cf. [41] [58]),

$$(6.6) \quad \sigma = \frac{\|J\|_2}{\int J' \circ F(x) f(x)^2 dx}$$

Sous cette forme le résultat ne suppose plus autant de régularité sur F que 6.1, 6.3 ; on peut espérer obtenir des théorèmes de normalité asymptotique de $\hat{\theta}_n$ avec des hypothèses fortes sur $J = \lim_{n \rightarrow \infty} J_n$ (ou $\varphi = \lim_{n \rightarrow \infty} \varphi_n$) et faibles sur F . C'est l'apport du théorème de CHERNOFF et SAVAGE [17] dans le cas de deux échantillons indépendants. Il est tentant "d'appliquer" ce théorème au cas des "deux" échantillons $-X_1 \dots -X_n$ et $X_1 \dots X_n$, et c'est ce que font HODGES et LEHMANN [41]. Mais nous avons ici stricte dépendance des échantillons ! La structure des démonstrations de CHERNOFF-SAVAGE reste valable, mais dans le détail technique, la majoration de certains termes "d'erreur" du second ordre sans l'aide de l'indépendance n'est pas évidente pour nous, (bien que plausible). Comme les auteurs de [2] [41] [58] utilisent 6.6 sans préciser les hypothèses nécessaires, nous présentons une version "minimale" du théorème de CHERNOFF-SAVAGE adaptée au cas étudié ici, avec hypothèses "maximales" sur J .

6.7. Théorème (d'après [17]) :

On suppose seulement F continue et symétrique. Soit J une fonction deux fois dérivables sur $]0[1[$; on suppose J'' (et donc J' , J) bornées, J croissante, et $J(u) + J(1-u) = 0$ pour $u \in [0,1]$. Alors le test de rang $h_n = \sum_{i=1}^n J\left(\frac{S_i}{2n+1}\right)$ - notations de 4.5 - est asymptotiquement normal sous P_u/\sqrt{n} pour tout $u \in \mathbb{R}$. On a

$$\lim_{n \rightarrow +\infty} P_{u/\sqrt{n}} \left\{ \frac{1}{\sigma_n} [h_n - \mu_n(u)] \leq a \right\} = \Phi(a) \quad u, a \in \mathbb{R}$$

avec

$$(6.8) \quad \begin{cases} \mu_n(u) = n \int J \left[\frac{1}{2} F(x) + \frac{1}{2} F\left(x + \frac{2u}{n}\right) \right] dF(x) \\ \sigma_n = \sqrt{n} \|J\|_2 \end{cases}$$

Preuve :

Fixons $u \in \mathbb{R}$. Posons $K_n(x) = F(x - u/\sqrt{n})$. Sous $P_{u/\sqrt{n}}$, $X_1 \dots X_n$ sont de même loi K_n . Soit $\tilde{K}_n(x)$ la proportion d'observations X_i dans $]-\infty, x]$ (distribution empirique). La distribution empirique $L_n(x)$ de $-X_1 \dots -X_n$ vérifie $\tilde{L}_n(x) = 1 - \tilde{K}_n(-x) + \epsilon_n(x)$ avec $\epsilon_n = 0$ ou $1/n$. La distribution empirique de l'échantillon symétrisé $-X_1 \dots -X_n$, $X_1 \dots X_n$ est $\tilde{H}_n = \frac{1}{2}(\tilde{K}_n + \tilde{L}_n)$. Pour n grand, \tilde{H}_n est proche de $H_n(x) = \frac{1}{2}[K_n(x) + 1 - K_n(-x)]$. On a clairement

$$I_n = \int J \circ \tilde{H}_n d\tilde{K}_n = \frac{1}{n} \sum_{i=1}^n J\left(\frac{S_i}{2n}\right)$$

et donc $|\sqrt{n} I_n - \frac{1}{\sqrt{n}} h_n| \leq \frac{\text{Cte}}{\sqrt{n}}$. Les hypothèses sur J donnent

$$J(\tilde{H}_n) = J(H_n) + (\tilde{H}_n - H_n) J'(H_n) + \delta_n |\tilde{H}_n - H_n|^2$$

où $|\delta_n| \leq \text{Cte}$. Par suite I_n s'écrit

$$I_n = \int J \circ H_n dK_n + \int J \circ H_n d(\tilde{K}_n - K_n) + \int (\tilde{H}_n - H_n) J' \circ H_n dK_n + R_n$$

où le "reste" R_n est donné par

$$(6.9) \quad R_n = \int (\tilde{H}_n - H_n) J' \circ H_n d(\tilde{K}_n - K_n) + \int \delta_n |\tilde{H}_n - H_n|^2 d\tilde{K}_n$$

Avec les notations 6.8, on a $\frac{1}{n} \mu_n(u) = \int J \circ H_n dK_n$. On montrera plus bas que pour tout $\epsilon > 0$, il existe $B > 0$ tel que

$$(6.10) \quad P_{u/\sqrt{n}} \{ |R_n| \leq B/n \} \geq 1 - \epsilon$$

Pour suite $(\sqrt{n} I_n - \frac{1}{\sqrt{n}} \mu_n(u))$ et donc $\frac{1}{\sqrt{n}} [h_n - \mu_n(u)]$ ont même loi limite que

$$D_n = \sqrt{n} \left[\int J \circ H_n d(\tilde{K}_n - K_n) + \int (\tilde{H}_n - H_n) J' \circ H_n dK_n \right]$$

Posons $\Gamma_n(x) = \int_0^x J' \circ H_n dK_n$; en intégrant par parties on obtient

$$\int (\tilde{H}_n - H_n) J' \circ H_n dK_n = - \int \Gamma_n d(\tilde{H}_n - H_n)$$

mais puisque

$$\tilde{H}_n(x) - H_n(x) = \frac{1}{2} [\tilde{K}_n(x) - K_n(x)] - \frac{1}{2} [\tilde{K}_n(-x) - K_n(-x)] + \frac{1}{2} \epsilon_n(x)$$

on voit alors que D'_n a même loi limite que

$$D'_n = \sqrt{n} \int Z_n d(\tilde{K}_n - K)$$

où
$$Z_n(x) = J \circ H_n(x) - \frac{1}{2} \Gamma_n(x) - \frac{1}{2} \Gamma_n(-x)$$

Par définition de \tilde{K}_n , K_n on a

$$D'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n [Z_n(X_i) - E\{Z_n(X_i)\}]$$

où E est l'espérance sous $P_{u/\sqrt{n}}$. Comme $|Z_n(x)| \leq \text{cte}$ pour tout $x \in \mathbb{R}$, tout n , les n variables aléatoires indépendantes $Y_{i,n} = Z_n(X_i) - E[Z_n(X_i)]$ sont uniformément bornées et le théorème limite central montre que D'_n est asymptotiquement normal $N(0, \sigma)$ avec $\sigma^2 = \lim_{n \rightarrow +\infty} [\text{var } Z_n(X_i)]$.

Le théorème de convergence dominée implique

$$\lim_{n \rightarrow +\infty} E[Z_n^\alpha(X_i)] = \lim_{n \rightarrow +\infty} \int Z_n^\alpha dK_n = \int Z^\alpha dF, \quad \alpha > 0$$

où $Z(x) = \lim_{n \rightarrow +\infty} Z_n(x + \frac{u}{\sqrt{n}})$. On vérifie immédiatement que $Z(x) = J \circ F(x)$ ce qui donne

$$\sigma^2 = \int (J \circ F)^2 dF - \left(\int J \circ F dF \right)^2 = \int_0^1 J^2(u) du$$

Avec les notations de (6.8) on voit que la loi de $\frac{1}{\sigma_n} [h_n - \mu_n(u)]$ sous $P_{u/\sqrt{n}}$ tend faiblement vers $N(0, 1)$. Il reste à prouver (6.10).

Soit $(U_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes définies sur un espace de probabilité (E, \mathcal{A}, P) , de loi uniforme sur $[0, 1]$. Soit G_n la distribution

empirique de $U_1 \dots U_n$. On sait que si $\Delta_n(x) = \sqrt{n} [G_n(x) - x]$, les variables aléatoires $\|\Delta_n\|_\infty = \sup_{0 \leq x \leq 1} |\Delta_n(x)|$ ont une loi limite (théorème de KOLMOGOROV [10]). En particulier pour tout $\epsilon > 0$ il existe B tel que

$$P \{ \|\Delta_n\|_\infty \leq B \} \geq 1 - \epsilon \quad \text{quel que soit } n.$$

Pour u et n fixés, la loi jointe des variables T_1, \dots, T_n où $T_i = K_n^{-1} \circ U_i$ est la même que celle de $X_1 \dots X_n$ sous $P_{u/\sqrt{n}}$. En particulier les lois de $[\tilde{K}_n(x) - K_n(x)]/\sqrt{n}$ et $\Delta_n \circ K_n(x)$ sous $P_{u/\sqrt{n}}$ et P resp. sont les mêmes. De (6.11) on tire donc

$$P_{u/\sqrt{n}} \left\{ \sup_{x \in \mathbb{R}} |\tilde{K}_n(x) - K_n(x)| \leq \frac{B}{\sqrt{n}} \right\} = P \{ \|\Delta_n\|_\infty \leq B \} \geq 1 - \epsilon$$

Mais sur l'ensemble $\Omega' = \{ \omega \mid \sup_{x \in \mathbb{R}} |\tilde{K}_n(x) - K_n(x)| \leq \frac{B}{\sqrt{n}} \}$ on a trivialement $|\tilde{H}_n - H_n| \leq \frac{2B+1}{\sqrt{n}}$, et par suite, d'après (6.9)

$$|R_n| \leq \frac{Cte}{n} B(2B+1) + \frac{Cte}{n} (2B+1)^2$$

ce qui prouve (6.10).

Q.E.D.

6.12. Corollaire : (notations et hypothèses du théorème 6.7) :

Supposons J croissante, $J(u) + J(1-u) \equiv 0$, J' , J'' bornées F continue et symétrique. Supposons que

$$\ell = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} \int J \left[\frac{1}{2} F(x) + \frac{1}{2} F(x + 2\alpha) \right] dF(x)$$

existe. Alors les estimateurs $\hat{\theta}_n$ associés aux tests de rang $h_n = \sum_{i=1}^n J \left(\frac{S_i}{2n+1} \right)$ sont asymptotiquement normaux et

$$\lim_{n \rightarrow +\infty} P_{\theta} \{ \sqrt{n} (\hat{\theta}_n - \theta) \leq u \} = \Phi \left(\frac{u}{\sigma} \right) \quad \text{pour tout } u \in \mathbb{R}$$

avec $\sigma = \frac{\|J\|_2}{\ell}$; voir (6.13) pour le calcul de ℓ .

6.13. Remarque :

Ce corollaire se déduit directement de théorème 6.7 et proposition 3.3. De plus, comme $\int J \circ F \, dF = 0$ on a

$$I(\alpha) = \frac{1}{\alpha} \int J \left[\frac{1}{2} F(x) + \frac{1}{2} F(x + 2\alpha) \right] dF(x) = \int \frac{F(x + 2\alpha) - F(x)}{2\alpha} J' \circ C_\alpha(x) dF(x)$$

où $F(x) < C_\alpha(x) < F(x + 2\alpha)$. D'après le lemme de FATOU on a

$$(6.14) \quad \liminf_{\alpha \rightarrow 0} I(\alpha) \geq \int_{\mathbb{R}} J' \circ F(x) f^2(x) \, dx$$

si F admet la densité f par rapport à la mesure de LEBESGUE.

Dans les "bons" cas, on a bien entendu - voir 6.6 -

$$(6.15) \quad \ell = \lim_{\alpha \rightarrow 0} I(\alpha) = \int_{\mathbb{R}} J' \circ F(x) f^2(x) \, dx$$

Par exemple (6.15) est vraie, si $f(x) = F'(x)$ existe pour tout x et si $f(x+h) \leq f_0(x')$ pour tout x , tout $h \in [0, \epsilon]$ avec $f_0 \in L_1(\mathbb{R}, dF)$.

Si on a seulement $F \ll \text{LEBESGUE}$, de densité f , il semble possible de déduire (?) de (6.14) 6.7 que

$$\limsup_{n \rightarrow +\infty} s_n \leq \frac{\|J\|_2}{\int J' \circ F(x) f^2(x) \, dx}$$

où $s_n^2 = n \operatorname{var}(\hat{\theta}_n)$, que $\hat{\theta}_n$ soit asymptotiquement normal ou pas, pourvu que J soit comme en 6.7 (cf. [40]).

Notons enfin que tous les estimateurs introduits dans la section 5 - sauf l'estimateur de BICKEL-HODGES - correspondent à des scores vérifiant les hypothèses des théorèmes 6.1 et 6.2. Les scores de l'estimateur de HODGES-LEHMANN vérifient en outre les hypothèses des théorèmes 6.7, 6.12. Quant à l'estimateur de BICKEL-HODGES il n'est pas asymptotiquement normal (cf. [7]).

7 - APPLICATION A L'ESTIMATEUR DE HODGES-LEHMANN.

Les $\hat{\theta}_n$ de HODGES-LEHMANN correspondent aux scores $J\left(\frac{i}{2n+1}\right)$ - relatifs à l'échantillon symétrisé - où $J(x) = 2x - 1$ $x \in [0, 1]$. D'après la section 6, si F est assez régulière (cf. 6.2 et 6.13) $\sqrt{n}(\hat{\theta}_n - \theta)$ est asymptotiquement normal $N(0, \sigma^2)$ où $\sigma^2 = \frac{1}{2\sqrt{3} \int f^2(x) dx}$. En fait σ^2 est sans doute (?) un majorant de

$\limsup_{n \rightarrow +\infty} \text{var } \sqrt{n}(\hat{\theta}_n - \theta)$ quelle que soit $F \ll \text{LEBESGUE}$ (cf. [40]). Soit

$\bar{\theta}_n = \frac{X_1 + \dots + X_n}{n}$. Si F est de variance finie s^2 , on a $\text{var}[\sqrt{n}(\bar{\theta}_n - \theta)] = s^2$ et donc l'efficacité asymptotique de $\hat{\theta}_n$ par rapport à $\bar{\theta}_n$ est

$$e = \frac{s^2}{\sigma^2} = 12 \frac{\left[\int f^2(x) dx \right]^2}{\int x^2 f(x) dx}$$

En particulier on peut montrer (cf. [40]) que $e \geq 0,834$ quelle que soit F de variance finie et de densité appartenant à $L_2(\mathbb{R})$. Lorsque F est normale on trouve $e = \frac{3}{\pi}$ (comme dans le cas de deux échantillons, cf. [40] [38]).

8 - RELATION AVEC LES COMBINAISONS LINÉAIRES DE STATISTIQUE D'ORDRE.

Soit $X_1 \dots X_n$ l'échantillon observé, $Y_1 \dots Y_n$ l'échantillon ordonné associé. Il n'est pas possible en général, si $\hat{\theta}_n$ est un estimateur du type construit plus haut, d'écrire $\hat{\theta}_n$ sous la forme $\sum_{i=1}^n \alpha(i, n) Y_i$ où les $\alpha(i, n)$ sont des poids fixés. Par exemple, pour l'estimateur H-L, ceci est vrai dès que $n \geq 4$. Cependant $\hat{\theta}_n$ semble réaliser une telle combinaison linéaire mais avec des poids "dépendant des observations". L'interprétation heuristique qui "justifie" cette assertion est due à HODGES [38] pour l'estimateur H-L.

Soient $b(i) \quad i = 1 \dots n$ des scores croissants définissant un test de rang $h = \sum_{i=1}^n b(R_i) \text{sign } X_i$. On définit les $d_i = b(i) - b(i-1)$, $i \geq 2$, $d_1 = b(1)$. Soient $U_1 \dots U_n$ des variables indépendantes, de loi uniforme sur $[0,1]$, et soit $V_1 \dots V_n$ l'échantillon ordonné associé. Les X_i étant de loi F (de densité f) on peut supposer que $X_i = F^{-1}(U_i)$ et donc $Y_i = F^{-1}(V_i) \simeq F^{-1}(\frac{i}{n+1})$ car $V_i \simeq E(V_i) = \frac{i}{n+1}$.

On a alors

$$Y_{i+k} - Y_i \simeq F^{-1}(\frac{i}{n+1}) - F^{-1}(\frac{i}{n+1}) \simeq \frac{k}{n} \frac{dF^{-1}}{dt}(\frac{i}{n+1}) \simeq \frac{k}{n f(Y_i)}$$

Par suite, si $M_{i,j} = \frac{1}{2}(Y_i + Y_j)$ on a

$$M_{n+1-i+k, i+k} \simeq M_{n+1-i, i} + \frac{k}{n f(Y_i)}$$

Soit $k(i)$ l'entier k tel que $M_{i+k, n+1-i+k}$ soit le plus proche possible de l'estimateur $\hat{\theta}$ associé à h . On a alors $M_{n+1-i+k(i), i+k(i)} \simeq \hat{\theta}$ d'où

$$(8.1) \quad k(i) \simeq n f(Y_i) [\hat{\theta} - M_{n+1-i, i}]$$

Rappelons que $\hat{\theta}$ est la médiane de ν où ν charge seulement les $M_{j,k}$ $k \leq j$, avec $\nu[M_{j,k}] = d_{1+j-k}$.

Si $k(i) > 0$, les nombres $M_{n+1-i+k, i+k}$, $0 \leq k \leq k(i)$ "sont" inférieurs à $\hat{\theta}$ et la masse de cet ensemble de points pour ν est

$\delta(i) \simeq \frac{1}{c} d_{1+(N+1-2i)} |k(i)|$. Si $k(i) < 0$ ces nombres sont supérieurs à $\hat{\theta}$ pour $k(i) \leq k \leq 0$, et la masse de cet ensemble pour ν est encore $\delta(i)$. Comme $\hat{\theta}$

est la médiane de ν on doit donc avoir $\sum_{k(i) > 0} \delta(i) \simeq \sum_{k(i) < 0} \delta(i)$ et par suite

$\sum_{1 \leq i \leq \frac{n}{2}} d_{n+2-2i} k(i) \simeq 0$. On obtient donc d'après (8.1)

$$(8.2) \quad \hat{\theta} \simeq \sum_{1 \leq i \leq \frac{n}{2}} \alpha_i \frac{Y_i + Y_{n+1-i}}{2} \quad \text{avec}$$

$$\alpha_i = \frac{1}{a} d_{n+2-2i} f(Y_i) \quad \text{et} \quad a = \sum_{1 \leq i \leq \frac{n}{2}} d_{n+2-2i} f(Y_i)$$

En particulier pour l'estimateur de HODGES-LEHMANN les poids α_i "sont" proportionnels à $f(Y_i)$ ce qui fait pressentir une certaine adaptativité de cet estimateur.

9 -RÉSULTATS NUMÉRIQUES POUR n PETIT.

Les résultats expérimentaux du séminaire ANDREWS - BICKEL ... [2] permettent une comparaison des performances des estimateurs $(H-L)_n$ (HODGES-LEHMANN) et $(M)_n$ (moyenne des observations) pour n observations ($n = 10, 20, 40$). Ce point est repris en détail plus bas dans l'exposé XV ; donnons cependant les principales conclusions :

- (i) la convergence pour $n \rightarrow +\infty$ de $n \operatorname{var} (H-L)_n$ vers sa limite est "rapide".
- (ii) la convergence de la loi de $(H-L)_n$ vers la loi normale Φ est "rapide" si F est voisine de Φ , et "plus lente mais raisonnable" si F quelconque.
- (iii) les performances de $(H-L)_n$ sont remarquablement meilleures que celles de $(M)_n$ dans toutes les situations autres que celle où F est extrêmement proche de Φ , et pratiquement identiques à celles de $(M)_n$ dans ces derniers cas.

Remarquons cependant que dans l'étude [2] , $(H-L)$ est le seul R-estimateur étudié ce qui explique le choix fait ici ; en particulier $(H-L)$ n'a pas été isolé de la famille des R-estimateurs à l'aide d'une propriété d'optimalité particulière.

10 - ALGORITHME DE CALCUL.

Etant donné un test de rang $h = \sum_{i=1}^n J(S_i)$ (les S_i sont les rangs des X_i dans l'échantillon symétrisé) nous décrivons d'après JAECKEL [58] un mode de calcul de l'estimateur associé $\hat{\theta}$. Remarquons d'abord qu'il est toujours possible d'ordonner un échantillon de taille n en un "temps" (= nombre d'opérations) $O(n \log n)$. Ceci se voit par récurrence en coupant l'échantillon en deux, en ordonnant séparément chaque partie, et en regroupant méthodiquement les deux moitiés déjà ordonnées. Posons

$$h_r = h(X_1 - r, \dots, X_n - r), \quad r \in \mathbb{R};$$

Une fois ordonné l'échantillon $\mathcal{S} = \{-X_1, \dots, -X_n, X_1, \dots, X_n\}$, le calcul de h_r pour r donné nécessite seulement un temps $O(n)$. En effet, comme \mathcal{S} est déjà ordonné, il est facile de voir que $2n$ comparaisons suffisent à ordonner $\mathcal{S}' = \{-(X_1 - r), \dots, -(X_n - r), (X_1 - r), \dots, (X_n - r)\}$ et on calcule alors h_r à partir des rangs S'_i des $(X_i - r)$ dans \mathcal{S}' .

Le calcul de $\hat{\theta}$ se fait par itération : on part de $\theta_1 \leq \theta_2$ tels que $\theta_1 \leq \hat{\theta} \leq \theta_2$, faciles à obtenir une fois \mathcal{S} ordonné (par exemple $\theta_1 = \inf_i X_i$, $\theta_2 = \sup_i X_i$, mais on peut souvent faire mieux. cf. exposé XI). On calcule

$$h_{r_1} \text{ pour } r_1 = \frac{\theta_1 + \theta_2}{2}.$$

Si $h_{r_1} < 0$ (resp. > 0) on remplace $[\theta_1, \theta_2]$ par $[\theta_1, r_1]$ (resp. $[r_1, \theta_2]$) puisque h_r est décroissante en r . Après $\log n$ itérations de ce type, on obtient un intervalle contenant θ et de taille $\simeq \frac{\theta_2 - \theta_1}{n}$. Chaque itération nécessite un temps $O(n)$; donc en un temps $O(n \log n)$ on a obtenu une approximation de $\hat{\theta}_n$ à $\frac{Cte}{n}$ près, ce qui est plus que suffisant puisque la précision de $\hat{\theta}_n$ est de l'ordre de $\frac{1}{\sqrt{n}}$.

Dans l'étude [2] le temps de machine à calculer nécessaire pour obtenir H-L

(avec $n = 40$) était de $1,510^{-3}$ seconde tandis que le temps de calcul de la moyenne était de 10^{-4} seconde.

Robert AZENCOTT

Mathématiques

Université Paris VII

2, Place Jussieu

75221 PARIS CEDEX 05.

CHAPITRE VI

L-ESTIMATEURS

(Combinaisons linéaires des statistiques d'ordre)

par

Catherine HUBER

1 - INTRODUCTION.

$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ désignent les statistiques d'ordre d'un n -échantillon d'une loi sur \mathbb{R} de fonction de répartition $F_\theta = F(\cdot - \theta)$ et F_n la fonction de répartition empirique associée à l'échantillon. Il s'agit d'estimer le paramètre θ de translation, F étant supposée connue, en utilisant un estimateur T_n qui soit une combinaison linéaire à coefficients réels c_{ni} , $i = 1, 2, \dots, n$ des $X_{(i)}$.

$$(1.1) \quad T_n = \sum_{i=1}^n c_{ni} X_{(i)}$$

Sous certaines conditions sur F et sur la suite $(c_{n1}, \dots, c_{nn})_{n \in \mathbb{N}}$ des coefficients, T_n est asymptotiquement normal.

Diverses démonstrations de cette normalité asymptotique ont été données (SCHORACK [82], HUBER [46], CHERNOFF et al. [16], BICKEL [6], MOORE [72]) sous des conditions variées sur les coefficients, d'autant plus fortes qu'on est moins exigeant pour F . Dans le cadre de notre étude, il sera naturel de ne pas demander à F de vérifier des conditions trop fortes. Les démonstrations les plus intéressantes

sont fondées sur la convergence faible de la suite des processus empiriques des quantiles vers un processus gaussien (SHORACK [82] , HUBER [46] , BICKEL [6]).

C'est une démonstration de ce type qui est donnée ici : le principe en est décrit au paragraphe 2 et la démonstration proprement dite est faite au paragraphe 3. Le paragraphe 4 concerne la détermination d'un estimateur asymptotiquement efficace dans la classe (1.1) quand F est connue.

2 - REPRÉSENTATION DE T_n COMME FONCTIONNELLE D'UN PROCESSUS EMPIRIQUE.

On sait que si $(U_i)_{i=1, \dots, n}$ est un n -échantillon de la loi uniforme sur $[0, 1]$, $X_{(i)}$ et $F^{-1}(U_{(i)})$ ont la même loi. De plus, si T_n désigne la fonction de répartition empirique associée au n -échantillon uniforme et obtenue par interpolation linéaire entre les points $(U_{(i)}, \frac{i}{n+1})$, $i = 0, 1, \dots, n, n+1$, avec la convention que $U_{(0)} = 0$ et $U_{(n+1)} = 1$, on sait que la suite $(V_n)_{n \in \mathbb{N}}$ des processus empiriques des quantiles

$$(2.1) \quad V_n(t) = \sqrt{n} (T_n^{-1}(t) - t) \quad t \in [0, 1]$$

converge faiblement vers le pont brownien (BILLINGSLEY [9]) dans $\mathcal{C}_{[0,1]}$, ensemble des fonctions continues sur $[0, 1]$, muni de la norme uniforme ; le pont brownien, qu'on notera V_0 , est un processus gaussien centré défini sur $[0, 1]$ et tel que $E[V_0(s) V_0(t)] = s \wedge t - st$.

La convergence de V_n vers V_0 a même lieu dans un sens plus fort (PYKE et SHORAK [77]) : Si δ est un réel fixé strictement compris entre 0 et $\frac{1}{2}$ et q la fonction définie sur $[0, 1]$ par $q(t) = t^\delta (1-t)^\delta$, on désigne par $\mathcal{C}_{[0,1]}^\delta$ le sous-espace de $\mathcal{C}_{[0,1]}$ des fonctions x telles que $\sup_{t \in [0,1]} \frac{|x(t)|}{q(t)}$ soit fini, et on munit $\mathcal{C}_{[0,1]}^\delta$ de la norme ρ_δ

$$(2.2) \quad \rho_\delta(x) = \inf \{ M \mid |x(t)| \leq M q(t) \quad t \in [0, 1] \} \quad x \in \mathcal{C}_{[0,1]}^\delta$$

Pour tout δ strictement inférieur à $\frac{1}{2}$, V_0 est presque sûrement à trajectoires dans $\mathcal{C}_{[0,1]}^\delta$. De plus, on peut construire (SKOROHOD [84], PYKE et SHORAK [77]) sur un même espace probabilisé (Ω, \mathcal{A}, P) une suite $(V'_n)_{n \in \mathbb{N}}$ de processus et un pont brownien V'_0 tels que, pour tout n , V'_n ait la même loi que V_n et V'_n converge presque sûrement vers V'_0 dans $\mathcal{C}_{[0,1]}^\delta$, pour tout δ de $]0, \frac{1}{2}[$.

Notons $(\nu_n)_{n \in \mathbb{N}}$ la suite des mesures

$$\nu_n = \sum_{i=1}^n c_{ni} \delta_{\frac{i}{n+1}}, \quad T'_n = \int_0^1 F^{-1}\left(t + \frac{1}{\sqrt{n}} V'_n(t)\right) d\nu_n(t)$$

et supposons que $(\nu_n)_{n \in \mathbb{N}}$ admette une limite ν au sens de la convergence vague des mesures sur $[0,1]$ muni de la tribu de ses boréliens.

Alors, si $m_n = \int_0^1 F^{-1}(t) d\nu_n(t)$, $m = \int_0^1 F^{-1}(t) d\nu(t)$ et si ℓ_n est la fonctionnelle définie (formellement pour l'instant) sur $\mathcal{C}^\delta_{[0,1]}$ par

$$(2.2) \quad \ell_n(x) = \sqrt{n} \int_0^1 \left[F^{-1}\left(t + \frac{x(t)}{\sqrt{n}}\right) - F^{-1}(t) \right] d\nu_n(t) \quad x \in \mathcal{C}^\delta_{[0,1]}$$

la loi de $\sqrt{n}(T_n - m)$ est la même que celle de $\sqrt{n}(T'_n - m) = \sqrt{n}(m_n - m) + \ell_n(V'_n)$.

Laissant dorénavant tomber les "primes", on écrira

$$(2.3) \quad \sqrt{n}(T_n - m) = \sqrt{n}(m_n - m) + \ell_n(V_n)$$

Pour démontrer la normalité asymptotique de la suite de variables aléatoires

$(\sqrt{n}(T_n - m))_{n \in \mathbb{N}}$, on utilisera le théorème suivant dû à PROKHOROV [76] :

Théorème 2.1 :

Soient Ω_1 et Ω_2 deux espaces métriques munis de leurs tribus respectives des boréliens \mathcal{B}_1 et \mathcal{B}_2 , $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures sur $(\Omega_1, \mathcal{B}_1)$ convergeant faiblement vers μ , $(\ell_n)_{n \in \mathbb{N}}$ une suite d'applications continues de Ω_1 dans Ω_2 convergeant vers une application ℓ .

Alors, si la convergence de ℓ_n vers ℓ a lieu uniformément sur les compacts de Ω_1 , la suite des mesures images de μ_n par ℓ_n , soit $\ell_n(\mu_n)$ converge faiblement vers $\ell(\mu)$, mesure image de μ par ℓ .

Sous certaines conditions de régularité de F et de $(\nu_n)_{n \in \mathbb{N}}$ qui seront explicitées au paragraphe suivant, la suite réelle $(\sqrt{n}(m_n - m))_{n \in \mathbb{N}}$ tend vers 0, les fonctionnelles ℓ_n définies en (2.2) sont continues et convergent uniformément sur les compacts de $\mathcal{C}^\delta_{[0,1]}$, vers une fonctionnelle ℓ telle que $\ell(V_0)$ soit gaussienne.

La normalité asymptotique de $\sqrt{n}(T_n - m)$ résulte alors du théorème (2.1) où Ω_1 est $\mathcal{C}^\delta_{[01]}$ muni de la norme ρ_δ , Ω_2 est \mathbb{R} muni de sa tribu des boréliens, μ_n est la loi sur $\mathcal{C}^\delta_{[01]}$ induite par V_n , μ la loi induite sur $\mathcal{C}^\delta_{[01]}$ par le pont brownien V_0 et ℓ_n la fonctionnelle définie en (2.2).

3 - NORMALITÉ ASYMPTOTIQUE.

On suppose qu'il existe une fonction mesurable h définie sur $[0,1]$, k points u_j , $j = 1, 2, \dots, k$, fixés dans $]0,1[$ et k réels fixés d_j , $j = 1, 2, \dots, k$ tels que ν_n s'écrive

$$(3.1) \quad \nu_n = \frac{1}{n} \sum_{i=1}^n h\left(\frac{i}{n+1}\right) \delta_{\frac{i}{n+1}} + \sum_{j=1}^k d_j \delta_{\frac{[nu_j] + 1}{n+1}}$$

où $[u]$ désigne, comme dans la suite, la partie entière de u . Si λ désigne la mesure de LEBESGUE sur $[0,1]$, la suite des mesures $(\nu_n)_{n \in \mathbb{N}}$ converge vaguement vers ν

$$(3.2) \quad \nu = h \lambda + \sum_{j=1}^k d_j \delta_{u_j}$$

On fait sur h et F les hypothèses suivantes qu'on appellera globalement (H) :

(H₁) F^{-1} est continûment dérivable aux points u_1, u_2, \dots, u_k

(H₂) Il existe un compact $[AB]$ de $]0,1[$ tel que

(i) Sur $[AB]$, h est à variation bornée et h et F^{-1} sont sans discontinuité commune.

(ii) Sur le complémentaire de $[AB]$ dans $]0,1[$, l'une ou l'autre des deux conditions suivantes est réalisée :

(ii1) h est nulle,

(ii2) F^{-1} et h sont dérivables et il existe des réels positifs C, C' et des réels b, β , avec $\beta + b < \frac{1}{2}$, tels que

$$(3.3) \quad \left| \frac{dF^{-1}(t)}{dt} \right| < \frac{C}{t^{1+b} (1-t)^{1+b}}$$

$$(3.4) \quad \left| \frac{dh(t)}{dt} \right| < \frac{C'}{t^{1+\beta} (1-t)^{1+\beta}}$$

Remarquons que l'hypothèse (H) a les conséquences suivantes :

a) Les majorations (3.3) et (3.4) entraînent l'existence de deux constantes D et D' telles que, pour tout t de $]0,1[$, on ait $|F^{-1}(t)| < \frac{D}{t^b(1-t)^b}$ et $|h(t)| < \frac{D'}{t^\beta(1-t)^\beta}$. Dans le cas où b est strictement positif, la condition (3.3) entraîne donc l'existence, pour la loi F , des moments d'ordre a strictement inférieur à $\frac{1}{b-1}$, car $E(|X|^a) = \int_0^1 |F^{-1}(t)|^a dt \leq D^a \int_0^1 \frac{dt}{t^{a(b-1)}(1-t)^{a(b-1)}}$ qui converge dès que $a(b-1)$ est strictement inférieur à 1.

b) Si on pose

$$(3.5) \quad dK(t) = h(t) dF^{-1}(t) + d \left[\sum_{j=1}^k d_j (F^{-1})'(u_j) \delta_{u_j} \right]$$

on définit une mesure signée sur $]0,1[$, bornée sur tout compact de $]0,1[$. De plus, l'expression (3.5) définit sur $]0,1[$ une fonction $K(t)$, à une constante additive près, qui a les propriétés suivantes :

$$\int_0^1 |K(t)| dt < \infty, \quad \int_0^1 K^2(t) dt < \infty, \quad \begin{matrix} t(1-t)K(t) \xrightarrow[t \rightarrow 0]{t \rightarrow 1} 0, & t(1-t)K^2(t) \xrightarrow[t \rightarrow 0]{t \rightarrow 1} 0. \end{matrix}$$

En effet, $|K(t)|$ est, au voisinage de $t=0$ majoré par une fonction de l'ordre de $t^{-(b+\beta)}$ et $b+\beta$ a été supposé strictement inférieur à $\frac{1}{2}$. De même, au voisinage de $t=1$, $|K(t)|$ est majoré par une fonction de l'ordre de $(1-t)^{-(b+\beta)}$.

On notera dans la suite

$$(3.6) \quad \sigma^2 = \int_0^1 K^2(t) dt - \left(\int_0^1 K(t) dt \right)^2$$

qui existe et est fini d'après les remarques précédentes.

c) La fonctionnelle

$$(3.7) \quad \ell(x) = \int_0^1 x(t) dK(t) \quad x \in \mathcal{C}^\delta_{[0,1]}$$

existe pour tout δ de $]b+\beta, \frac{1}{2}[$. C'est clair en ce qui concerne la partie discontinue de la mesure K . Supposons donc que $d_1 = d_2 = \dots = d_k = 0$.

$$\text{Alors} \quad \int_0^1 |x(t)| dK(t) \leq L \int_0^1 t^\delta (1-t)^\delta \frac{1}{t^{b+\beta+1}(1-t)^{b+\beta+1}} dt$$

pour une constante positive L et l'intégrale de droite converge pour tout δ strictement supérieur à $b + \beta$.

On notera S la variable suivante

$$(3.8) \quad S = \ell(V_O)$$

Théorème 3.1:

Sous l'hypothèse (H), $\sqrt{n}(T_n - m)$ converge en loi vers $S = \ell(V_O)$, variable aléatoire gaussienne centrée de variance $\sigma^2 = \int_0^1 K^2(t) - \left(\int_0^1 K(t) dt\right)^2$.

Démonstration:

La variable aléatoire $\sqrt{n}(T_n - m)$ est la somme de deux termes (2.3) dont nous allons d'abord montrer que le premier tend vers 0.

Etude de $\sqrt{n}(m_n - m)$

$$\begin{aligned} \sqrt{n}(m_n - m) &\leq \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n h\left(\frac{i}{n+1}\right) F^{-1}\left(\frac{i}{n+1}\right) - \int_0^1 F^{-1}(t) h(t) dt \right| \\ &\quad + \sqrt{n} \sum_{j=1}^k |d_j| \left| F^{-1}\left(\frac{[nu_j] + 1}{n+1}\right) - F^{-1}(u_j) \right| \\ &= \gamma_{n1} + \gamma_{n2} \end{aligned}$$

Comme F^{-1} a été supposée continûment dérivable en chaque point u_j , $j = 1, 2, \dots, k$ (H_1), le deuxième terme γ_{n2} tend vers 0.

Pour démontrer que γ_{n1} tend vers zéro, considérons séparément ce qui se passe sur le compact $[AB]$ et sur $]0A[$ (le traitement de $]B1[$ serait analogue à celui de $]0A[$) en notant $h^{(1)}(t) = h(t)1_{[AB]}(t)$ et $h^{(2)}(t) = h(t)1_{]0A[}(t)$.

$$\text{Soit} \quad \gamma'_{n1} = \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n h^{(1)}\left(\frac{i}{n+1}\right) F^{-1}\left(\frac{i}{n+1}\right) - \int_0^1 F^{-1}(t) h^{(1)}(t) dt \right|$$

Il suffit de démontrer que γ'_{n1} tend vers 0 lorsque $h^{(1)}$ est positive croissante et bornée sur $[AB]$ puisque γ'_{n1} est linéaire en $h^{(1)}$ et h est à variation bornée sur $[AB]$; supposons donc $h^{(1)}$ positive croissante et bornée :

$$\begin{aligned} \sqrt{n} m_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n h^{(1)} \left(\frac{i}{n+1} \right) F^{-1} \left(\frac{i}{n+1} \right) \leq \sqrt{n} m \leq \frac{1}{\sqrt{n}} \sum_{i=1}^n h^{(1)} \left(\frac{i+1}{n+1} \right) F^{-1} \left(\frac{i+1}{n+1} \right) \\ &\leq \sqrt{n} m_n + \frac{1}{\sqrt{n}} [h^{(1)}(B) F^{-1}(B) - h^{(1)}(A) F^{-1}(A)] \end{aligned}$$

Donc $|\sqrt{n}(m_n - m)|$, qui est majoré par $\frac{1}{\sqrt{n}} [h^{(1)}(B) F^{-1}(B) - h^{(1)}(A) F^{-1}(A)]$ tend vers 0 quand n tend vers l'infini.

En ce qui concerne l'extérieur du compact $[AB]$, soit γ''_{n1} le terme correspondant à $]0A[$:

$$\begin{aligned} \gamma''_{n1} &= \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n h^{(2)} \left(\frac{i}{n+1} \right) F^{-1} \left(\frac{i}{n+1} \right) - \int_0^1 F^{-1}(t) h^{(2)}(t) dt \right| \\ &\leq \sqrt{n} \int_0^{\frac{1}{n}} |F^{-1}(t) h(t)| dt + \frac{[nA]-1}{\sum_{i=1}^n} \sqrt{n} \int_{\frac{i}{n}}^{\frac{i+1}{n}} |F^{-1}(t) h(t) - F^{-1} \left(\frac{i}{n+1} \right) h \left(\frac{i}{n+1} \right)| dt \\ &\quad + \sqrt{n} \int_{\frac{[nA]+1}{n}}^A |F^{-1}(t) h(t)| dt \\ &= \sqrt{n} \xi(n) + \sqrt{n} \xi'(n) + \sqrt{n} \xi''(n) \end{aligned}$$

D'une part, $\xi(n) = o\left(\frac{1}{\sqrt{n}}\right)$ car l'hypothèse (H_2) entraîne l'existence d'un δ' strictement positif tel que $\int_0^u F^{-1}(t) h(t) dt = o(u^{\frac{1}{2} + \delta'})$, et $\xi''(n)$ est évidemment $o\left(\frac{1}{n}\right)$.

D'autre part, $F^{-1}h$ étant dérivable sur $]0A[$, on peut majorer, pour $i = 1, \dots, [nA] + 1$, la différence $|F^{-1}(t) h(t) - F^{-1} \left(\frac{i}{n+1} \right) h \left(\frac{i}{n+1} \right)|$ par $L \left(\frac{i+1}{n+1} \right)^{-(b+\beta)}$ pour une constante positive L (3.3), (3.4), remarque a)). Par suite $\xi'(n)$ a le même comportement asymptotique que l'intégrale $\int_{\frac{1}{n}}^A \frac{dt}{t^{b+\beta}}$ qui est un $o\left(\frac{1}{\sqrt{n}}\right)$.

Donc γ''_{n1} tend vers 0 quand n croît, et comme on a vu que γ'_{n1} tend aussi vers 0, cela achève de démontrer que $\sqrt{n}(m_n - m)$ tend vers 0 quand n croît.

Etude de $\ell_n(V_n)$.

Il s'agit de démontrer que les fonctionnelles ℓ_n sont continues et que la suite de fonctionnelles $(\ell_n)_{n \in \mathbb{N}}$ (2.2) converge vers ℓ (3.7) uniformément sur les compacts de $\mathcal{C}^\delta_{[01]}$ si δ est strictement compris entre $b + \beta$ et $\frac{1}{2}$.

Dans la définition de ν_n (3.1), on peut, sans restreindre la généralité de la démonstration, supposer, pour simplifier l'écriture, que $k = 1$ et noter $d_1 = d$ et $u_1 = u$.

Commençons par démontrer que les fonctionnelles ℓ_n sont continues sur $\mathcal{C}^\delta_{[01]}$. Soit donc n fixé dans \mathbb{N} et x un point fixé dans $\mathcal{C}^\delta_{[01]}$. Si y est un autre point de $\mathcal{C}^\delta_{[01]}$

$$\begin{aligned} \ell_n(x) - \ell_n(y) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[F^{-1} \left(\frac{i}{n+1} + \frac{x \left(\frac{i}{n+1} \right)}{\sqrt{n}} \right) - F^{-1} \left(\frac{i}{n+1} + \frac{y \left(\frac{i}{n+1} \right)}{\sqrt{n}} \right) \right] h \left(\frac{i}{n+1} \right) \\ &\quad + \sqrt{n} d \left[F^{-1} \left(\frac{\lfloor nu \rfloor + 1}{n+1} + \frac{x \left(\frac{\lfloor nu \rfloor + 1}{n+1} \right)}{\sqrt{n}} \right) \right. \\ &\quad \left. - F^{-1} \left(\frac{\lfloor nu \rfloor + 1}{n+1} + \frac{y \left(\frac{\lfloor nu \rfloor + 1}{n+1} \right)}{\sqrt{n}} \right) \right] \\ &= \xi_1(y) + \xi_2(y) \end{aligned}$$

Dans le cas où aucun des n points (fixés) $\frac{i}{n+1} + \frac{x \left(\frac{i}{n+1} \right)}{\sqrt{n}}$, $i = 1, 2, \dots, n$, n'est un point de discontinuité de F^{-1} , il est clair que $|\xi_1(y)|$ tend vers 0 quand $\rho_\delta(x, y)$ tend vers 0.

Soit E_n l'ensemble des points x de $\mathcal{C}^\delta_{[01]}$ tels que F^{-1} soit discontinue en au moins l'un des n points $\frac{i}{n+1} + \frac{x \left(\frac{i}{n+1} \right)}{\sqrt{n}}$, $i = 1, 2, \dots, n$. Cet ensemble est de mesure nulle pour les lois μ_n et μ induites respectivement par V_n et par V_0 .

sur $\mathcal{C}^\delta_{[01]}$. Par suite $\bigcup_{n \in \mathbb{N}} E_n$ est lui aussi négligeable pour la mesure μ .

En ce qui concerne $\xi_2(y)$, on peut faire le même raisonnement : Sauf pour un ensemble μ -négligeable E'_n de $\mathcal{C}^\delta_{[01]}$, égal à l'ensemble des x de $\mathcal{C}^\delta_{[01]}$ tels que F^{-1} soit discontinue au point $\frac{\llbracket nu \rrbracket + 1}{n+1} + \frac{1}{\sqrt{n}} x (\frac{\llbracket nu \rrbracket + 1}{n+1})$, $|\xi_2(y)|$ tend vers 0 quand $\rho_\delta(x, y)$ tend vers 0.

Les fonctionnelles ℓ_n sont donc presque sûrement continues sur $\mathcal{C}^\delta_{[01]}$, pour la mesure induite par V_0 . Les compacts de $\mathcal{C}^\delta_{[01]}$ sont constitués par les sous-ensembles bornés et équicontinus de $\mathcal{C}^\delta_{[01]}$, c'est-à-dire que si K est un tel compact, d'une part, il existe M dans \mathbb{R}^+ tel que, pour tout t de $\llbracket 01 \rrbracket$, $|x(t)| < M t^\delta (1-t)^\delta$ pour tout x de K , d'autre part, pour tout ϵ positif, il existe η positif tel que $|t' - t| < \eta$ entraîne $|x(t) - x(t')| < \epsilon$ pour tout x de K .

$$\begin{aligned} \ell_n(x) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[F^{-1} \left(\frac{i}{n+1} + \frac{x(\frac{i}{n+1})}{\sqrt{n}} \right) - F^{-1} \left(\frac{i}{n+1} \right) \right] h \left(\frac{i}{n+1} \right) \\ &\quad + \sqrt{n} d \left[F^{-1} \left(\frac{\llbracket nu \rrbracket + 1}{n+1} + \frac{x(\frac{\llbracket nu \rrbracket + 1}{n+1})}{\sqrt{n}} \right) - F^{-1} \left(\frac{\llbracket nu \rrbracket + 1}{n+1} \right) \right] \\ &= \ell'_n(x) + \ell''_n(x) \end{aligned}$$

Comme F^{-1} a été supposée continûment dérivable en u , $\ell''_n(x)$, qui vaut $(F^{-1})'(u + O(\frac{1}{\sqrt{n}})) \times x(u + O(\frac{1}{\sqrt{n}}))$ converge vers $(F^{-1})'(u) x(u)$ uniformément sur K .

Considérons maintenant ℓ'_n .

Soit $\epsilon > 0$ fixé. On peut toujours trouver un découpage fini

$(t_0 = 0, t_1, \dots, t_m, t_{m+1} = 1)$ de $[01]$ vérifiant les conditions suivantes

(1) Les t_j sont des points de continuité de F^{-1} ;

(2) Les extrémités t_1 et t_m du découpage sont choisies de telle sorte que

$$\int_0^{t_1} t^{\delta-(1+b+\beta)} dt < \epsilon \quad \text{et} \quad \int_{t_m}^1 (1-t)^{\delta-(1+b+\beta)} dt < \epsilon$$

L'hypothèse (H) entraîne que sur tout compact de $]0,1[$ et donc en particulier sur le segment $[t_1, t_m]$, les fonctions h , h_1 et h_2 , pour tout x de $\mathcal{C}^\delta_{[0,1]}$, sont intégrables au sens de RIEMANN par rapport à dF^{-1} . D'autre part, h étant, par hypothèse, à variation bornée sur $[t_1, t_m]$ peut être considérée comme la différence de deux fonctions positives croissantes et bornées sur $[t_1, t_m]$, soit $h(t) = h_1(t) - h_2(t)$ $t \in [t_1, t_m]$. On notera h_r , $r = 1, 2$ ces fonctions.

Les points intermédiaires t_2, \dots, t_{m-1} du découpage seront choisis de telle sorte que certaines sommes de RIEMANN approchent les intégrales correspondantes, sur $[t_1, t_m]$:

(3) Pour tout x du compact K de $\mathcal{C}^\delta_{[0,1]}$ et pour $r = 1, 2$

$$\left| \int_{t_1}^{t_m} h_r(t) x(t) dF^{-1}(t) - \sum_{j=1}^{m-1} h_r(t_j) x(t_j) [F^{-1}(t_{j+1}) - F^{-1}(t_j)] \right| < \epsilon$$

la possibilité d'avoir l'uniformité en K résulte de ce que les éléments de K sont uniformément équicontinus.

$$(4) \quad \left| \int_{t_1}^{t_m} h_r(t) dF^{-1}(t) - \sum_{j=1}^{m-1} h_r(t_j) [F^{-1}(t_{j+1}) - F^{-1}(t_j)] \right| < \epsilon$$

pour $r = 1, 2$

(5) Notant $q_j = \sup_{t \in [t_j, t_{j+1}]} q(t)$, les points (t_j) , $j = 2, \dots, m-1$ sont choisis de telle sorte qu'on ait à la fois

$$(5') \quad \left| \int_{t_1}^{t_m} q(t) dh(t) - \sum_{j=1}^{m-1} [h(t_{j+1}) - h(t_j)] q_j \right| < \epsilon$$

$$(5'') \quad \left| \sum_{j=1}^{m-1} [h(t_{j+1}) - h(t_j)] [F^{-1}(t_{j+1}) - F^{-1}(t_j)] q_j \right| < \epsilon$$

Il est possible d'obtenir, par un découpage convenable, l'inégalité (5') parce que q , qui est continue, est intégrable par rapport à dh sur $[t_1, t_m]$, et l'inégalité (5'') parce que le membre de droite est la valeur absolue de la différence de deux sommes de RIEMANN qui convergent toutes les deux vers

$$\int_{t_1}^{t_m} h(t) q(t) dF^{-1}(t) .$$

On demande enfin à (t_1, \dots, t_m) d'avoir la propriété suivante :

(6) Pour tout $j = 1, 2, \dots, m-1$, on a $|t_{j+1} - t_j| < \eta$, si η est un réel positif tel que, pour tout (t, t') de $[0, 1]^2$ vérifiant $|t - t'| < \eta$, on ait $|x(t) - x(t')| < \epsilon$.

Dans la suite, le découpage, fonction de ϵ , sera supposé fixé et on fera tendre n vers l'infini.

Notons

$$\alpha_i(n) = \frac{1}{\sqrt{n}} x\left(\frac{i}{n+1}\right)$$

$$\beta_j(n) = \frac{1}{n+1} \llbracket \sqrt{n} x(t_j) \rrbracket$$

$$I_j(n) = \{i \in \mathbb{N} \mid t_j \leq \frac{i}{n+1} < t_{j+1}\}$$

$$A_j(n) = h(t_j) \sum_{i \in I_j} [F^{-1}\left(\frac{i}{n+1} + \beta_j(n)\right) - F^{-1}\left(\frac{i}{n+1}\right)]$$

$$B_j(n) = h(t_j) \sum_{i \in I_j} [F^{-1}\left(\frac{i}{n+1} + \alpha_i(n)\right) - F^{-1}\left(\frac{i}{n+1} + \beta_j(n)\right)]$$

$$C_j(n) = \sum_{i \in I_j} \left[h\left(\frac{i}{n+1}\right) - h(t_j)\right] [F^{-1}\left(\frac{i}{n+1} + \alpha_i(n)\right) - F^{-1}\left(\frac{i}{n+1}\right)]$$

$$D(n) = \frac{1}{\sqrt{n}} \sum_{i \in I_0 \cup I_m} h\left(\frac{i}{n+1}\right) [F^{-1}\left(\frac{i}{n+1} + \alpha_i(n)\right) - F^{-1}\left(\frac{i}{n+1}\right)]$$

Avec ces notations $\ell'_n(x)$ peut s'écrire

$$(3.9) \quad \ell'_n(x) = \frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} (A_j(n) + B_j(n) + C_j(n)) + D(n)$$

On étudiera successivement les comportements de $D(n)$, $A_j(n)$, $B_j(n)$ et $C_j(n)$.

Etude de $D(n)$:

Il suffit de regarder ce qui se passe pour i variant dans I_0 , l'autre terme ayant un comportement analogue. On omettra les signes "partie entière" $\llbracket \quad \rrbracket$ chaque fois que cela ne peut pas créer de confusion.

Il existe une constante positive L et un réel strictement positif δ' tels que

$$\begin{aligned} & \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{(n+1)t_1} h\left(\frac{i}{n+1}\right) \left[F^{-1}\left(\frac{i}{n+1} + \alpha_i(n)\right) - F^{-1}\left(\frac{i}{n+1}\right) \right] \right| \\ & \leq \frac{L}{n} \sum_{i=1}^{(n+1)t_1} \left(\frac{i}{n+1}\right)^{\delta - (1+b+\beta)} \left(1 + o\left(\frac{1}{n^{\delta'}}\right)\right) \end{aligned}$$

Comme, lorsque n tend vers l'infini, la somme $\frac{1}{n} \sum_{i=1}^{(n+1)t_1} \left(\frac{i}{n+1}\right)^{\delta - (1+b+\beta)}$ converge vers l'intégrale $\int_0^{t_1} t^{\delta - (1+b+\beta)} dt$ qui, d'après la propriété (2) du découpage, est inférieure à ϵ , $D(n)$ peut être rendu arbitrairement proche de ϵ pour n assez grand, soit $D(n) < 2\epsilon$ dès que n est supérieur à N_0 .

Etude de $A_j(n)$:

Les termes centraux s'annulent deux à deux et il ne reste que ceux qui sont proches de t_j et de t_{j+1} , en nombre de l'ordre de \sqrt{n} :

$$A_j(n) = h(t_j) \left[\sum_{t_{j+1} \leq \frac{i}{n+1} < t_{j+1} + \beta_j(n)} F^{-1}\left(\frac{i}{n+1}\right) - \sum_{t_j \leq \frac{i}{n+1} < t_j + \beta_j(n)} F^{-1}\left(\frac{i}{n+1}\right) \right]$$

Par définition, $(n+1)\beta_j(n) = \llbracket \sqrt{n} x(t_j) \rrbracket = \sqrt{n} x(t_j) + O(1)$. Donc, sur le compact K , $\beta_j(n) = O\left(\frac{1}{\sqrt{n}}\right)$ et comme F^{-1} est continue en chaque point t_j , $j = 1, 2, \dots, m$ (propriété (1) du découpage) $\frac{1}{\sqrt{n}} A_j(n)$ converge, quand n croît, vers $h(t_j) [F^{-1}(t_{j+1}) - F^{-1}(t_j)] x(t_j)$. D'après la propriété (3) du découpage, la somme $\frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} A_j(n)$ peut donc être rendue proche, à moins de 2ϵ , de l'intégrale $\int_{t_1}^m h(t) x(t) dF^{-1}(t)$, dès que n est supérieur à N_1 . Comme la propriété (2) du découpage entraîne qu'il existe une constante L telle que, pour tout x de K

$$\left| \int_0^1 h(t) x(t) dF^{-1}(t) - \int_{t_1}^{t_m} h(t) x(t) dF^{-1}(t) \right| < L\epsilon$$

il s'ensuit que, pour tout n supérieur à N_1

$$\left| \frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} A_j(n) - \int_0^1 h(t) x(t) dF^{-1}(t) \right| < \epsilon(L+2)$$

uniformément sur le compact K .

Etude de $B_j(n)$:

La fonction h est à variation bornée sur le compact $[t_1, t_m]$ et $B_j(n)$ est linéaire en h , on peut donc supposer h croissante positive et bornée pour étudier la limite de $B_j(n)$, c'est-à-dire remplacer h par h_r , $r = 1, 2$ dans $B_j(n)$:

$$h(t) = h_1(t) - h_2(t), \quad t \in [t_1, t_m].$$

D'après la propriété (6) du découpage, pour tout $j = 1, 2, \dots, m-1$:

$$\alpha_i(n) \leq \frac{1}{\sqrt{n}} (x(t_j) + \epsilon) \leq \frac{1}{n+1} \{ \lfloor \frac{n+1}{\sqrt{n}} (x(t_j) + \epsilon) \rfloor + 1 \} = \gamma_j(n)$$

Comme h_r est positive et comme F^{-1} est croissante

$$\begin{aligned} |B_j(n)| &\leq h_r(t_j) \sum_{i \in I_j} [F^{-1}(\frac{i}{n+1} + \gamma_j(n)) - F^{-1}(\frac{i}{n+1} + \beta_j(n))] \\ &= h_r(t_j) \left[\sum_{t_{j+1} + \beta_j(n) \leq \frac{i}{n+1} < t_{j+1} + \gamma_j(n)} F^{-1}(\frac{i}{n+1}) - \sum_{t_j + \beta_j(n) \leq \frac{i}{n+1} < t_j + \gamma_j(n)} F^{-1}(\frac{i}{n+1}) \right] \end{aligned}$$

Or, les t_j sont des points de continuité de F^{-1} et $\gamma_j(n)$ et $\beta_j(n)$ sont des $O(\frac{1}{\sqrt{n}})$ uniformément sur le compact K . En remarquant que le nombre $(n+1)(\gamma_j(n) - \beta_j(n))$ des termes de chacune des sommes entre crochets est majoré par $L\sqrt{n}\epsilon$, pour une constante positive L , il vient

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} B_j(n) \leq \left(\sum_{j=1}^{m-1} h_r(t_j) [F^{-1}(t_{j+1}) - F^{-1}(t_j)] \right) L\epsilon$$

D'après la propriété (4) du découpage, $\sum_{j=1}^{m-1} h(t_j) [F^{-1}(t_{j+1}) - F^{-1}(t_j)]$ est proche à ϵ de l'intégrale $\int_{t_1}^{t_m} h(t) dF^{-1}(t) = J$, nombre fini.

Par suite, il existe N_2 tel que, pour tout n supérieur à N_2

$$\left| \frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} B_j(n) \right| \leq (|J| + \epsilon) L\epsilon.$$

Etude de $C_j(n)$:

On va démontrer que $\frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} C_j(n)$ tend, quand n croît, vers un $O(\epsilon)$ en sup-

posant, comme lors de l'étude de $B_j(n)$, que h est remplacée par h_r positive croissante et bornée sur le compact $[t_1, t_m]$. Posons

$$\Delta_j(n) = \frac{1}{n+1} \{ \lfloor \frac{n+1}{\sqrt{n}} M q_j \rfloor + 1 \}$$

Alors, on a

$$\begin{aligned} |C_j(n)| &\leq (h_r(t_{j+1}) - h_r(t_j)) \sum_{i \in I_j} [F^{-1}(\frac{i}{n+1} + \Delta_j(n)) - F^{-1}(\frac{i}{n+1})] \\ &= (h_r(t_{j+1}) - h_r(t_j)) [\sum_{t_{j+1} \leq \frac{i}{n+1} < t_{j+1} + \Delta_j(n)} F^{-1}(\frac{i}{n+1}) \\ &\quad - \sum_{t_j \leq \frac{i}{n+1} < t_j + \Delta_j(n)} F^{-1}(\frac{i}{n+1})] \end{aligned}$$

D'une part, il existe une constante M' telle que $(n+1) \Delta_j(n) < M' \sqrt{n} q_j$, pour tout $j = 1, 2, \dots, m-1$. D'autre part, si on note

$$R(n, \epsilon) = \max_{j=1, \dots, m} ((F^{-1}(t_{j+1} + \Delta_j(n)) - F^{-1}(t_{j+1})) ,$$

comme ϵ et (t_1, \dots, t_m) sont fixés, $R(n, \epsilon) \xrightarrow{n \rightarrow \infty} 0$. Par suite

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} |C_j(n)| \leq M' \sum_{j=1}^{m-1} [h_r(t_{j+1}) - h_r(t_j)] [F^{-1}(t_{j+1}) - F^{-1}(t_j) + R(n, \epsilon)] q_j$$

$$\leq M' \epsilon + M' (|J_1| + \epsilon) R(n, \epsilon)$$

si on désigne par J_1 l'intégrale $\int_{t_1}^{t_m} q(t) d h_r(t)$, et d'après les propriétés (5') et (5'') du découpage.

On a donc démontré que lorsque n tend vers l'infini $\frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} (B_j(n) + C_j(n)) + D(n)$ converge vers un $O(\epsilon)$ et que $\frac{1}{\sqrt{n}} \sum_{j=1}^{m-1} A_j(n)$ converge vers $\int_0^1 x(t) h(t) dF^{-1}(t)$ à un $O(\epsilon)$ près.

En faisant tendre maintenant ϵ vers 0, on obtient le résultat cherché

$$\ell'_n(x) \xrightarrow{n \rightarrow \infty} \ell(x)$$

uniformément sur les compacts de $\mathcal{C}_0^\delta[0,1]$.

Pour achever la démonstration du théorème, il reste à prouver que la variable $S = \ell(V_0)$ définie en (3.7) et (3.8) est gaussienne centrée de variance $\sigma^2 = \int_0^1 K^2(t) dt - \left(\int_0^1 K(t) dt \right)^2$. Rappelons que K est défini en (3.5).

Le fait que S est gaussienne et centrée résulte de ce que ℓ est une fonctionnelle linéaire d'un processus gaussien centré, le pont brownien V_0 .

Calculons maintenant la variance de S :

$$\begin{aligned} \text{Var } S &= E \left[\int_0^1 \int_0^1 V_0(s) V_0(t) dK(s) dK(t) \right] \\ &= \int_0^1 \int_0^1 (s \wedge t - st) dK(s) dK(t) \end{aligned}$$

puisque cette dernière intégrale est absolument convergente d'après l'hypothèse (H) (remarque b)). On peut donc écrire

$$\text{Var } (S) = \lim_{\substack{u \rightarrow 0 \\ v \rightarrow 1}} 2 \int_u^v \left(\int_u^t s(1-t) dK(s) \right) dK(t) = 2 \lim I(u, v)$$

La remarque b) qui suit l'hypothèse (H) justifie le calcul suivant de $I(u, v)$ par intégrations par parties successives :

$$\begin{aligned} I(u, v) &= \int_u^v \{ t K(t) - u K(u) - \int_u^t K(s) ds \} (1-t) dK(t) \\ &= \left[\frac{1}{2} K^2(t) t (1-t) \right]_u^v - \int_u^v \frac{K^2(t)}{2} d(t(1-t)) - \int_u^v \left(\int_s^v (1-t) dK(t) \right) K(s) ds \\ &\quad + o(u) \\ &= -\frac{1}{2} \int_u^v K^2(t) (1-2t) dt - \int_u^v \{ [K(v)(1-v) - K(s)(1-s)] + \int_s^v K(t) dt \} K(s) ds \\ &\quad + o(u) + o(1-v) \\ &= -\frac{1}{2} \int_u^v K^2(t) (1-2t) dt + \int_u^v K^2(s) (1-s) ds - \int_u^v \left(\int_s^v K(t) dt \right) K(s) ds \\ &\quad + o(u) + o(1-v) \end{aligned}$$

Si on note $H(t) = \int_a^t K(s) ds$, a réel quelconque de $]0,1[$, une primitive de K , on a finalement

$$\begin{aligned} I(u, v) &= o(u) + o(1-v) + \frac{1}{2} \int_u^v K^2(t) dt - H(v)(H(v) - H(u)) + \frac{H^2(v)}{2} - \frac{H^2(u)}{2} \\ &= o(u) + o(1-v) + \frac{1}{2} \int_u^v K^2(t) dt - \frac{1}{2} (H(v) - H(u))^2 \\ &\xrightarrow[u \rightarrow 0]{v \rightarrow 1} \frac{1}{2} \int_0^1 K^2(t) dt - \frac{1}{2} \left(\int_0^1 K(t) dt \right)^2 \end{aligned}$$

Remarque :

On a souvent besoin de considérer une fonction de poids h qui varie avec n , soit $h = h_n$ et $d_j = d_{jn}$, $j = 1, 2, \dots, k$; par exemple, pour la moyenne α tronquée,

$$h_n(t) = \frac{n}{n - 2 \llbracket n\alpha \rrbracket}. \quad \text{Notons}$$

$$(3.10) \quad T_n^* = \frac{1}{n} \sum_{i=1}^n h_n\left(\frac{i}{n+1}\right) X_{(i)} + \sum_{j=1}^k d_{jn} X_{\llbracket n u_j \rrbracket} + 1$$

On appellera (H^*) l'hypothèse suivante :

(H_1^*) Pour tout n , le couple (h_n, F) vérifie l'hypothèse (H) pour un compact $[AB]$ de $]0,1[$ et des constantes, C, C', b, β .

(H_2^*) Il existe une fonction h définie sur $]0,1[$ et telle que, d'une part le couple (h, F) vérifie l'hypothèse (H) pour le compact $[AB]$ et les mêmes constantes et, d'autre part h_n tende vers h de telle sorte que sur tout compact D de $]0,1[$

$$\sup_{t \in D} |h_n(t) - h(t)| = o\left(\frac{1}{\sqrt{n}}\right)$$

(H_3^*) Il existe des réels d_1, \dots, d_k tels que $\sqrt{n}(d_{jn} - d_j)$ tende vers 0 quel que soit $j = 1, 2, \dots, k$.

Le théorème (3.1) admet le corollaire suivant où m vaut $\int_0^1 F^{-1}(t) d\nu(t)$ et ν et K sont définis en (3.2) et (3.5).

Corollaire 3.1 :

Sous l'hypothèse (H^*) , la suite de variables aléatoires $\sqrt{n}(T_n^* - m)$ converge en loi vers une variable gaussienne centrée de variance égale à

$$\sigma^2 = \int_0^1 K^2(t) dt - \left(\int_0^1 K(t) dt \right)^2.$$

Démonstration :

Le corollaire sera une conséquence immédiate du théorème (3.1) si on démontre que $\sqrt{n}(T_n^* - T_n)$ tend vers 0 en probabilité.

$$\begin{aligned} \sqrt{n}(T_n^* - T_n) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(h_n\left(\frac{i}{n+1}\right) - h\left(\frac{i}{n+1}\right) \right) X_{(i)} + \sqrt{n} \sum_{j=1}^k (d_{jn} - d_j) X_{[n u_j]} + 1 \\ &= \xi_1(n) + \xi_2(n) \end{aligned}$$

Comme $X_{[n u_j]} + 1 = F^{-1}(u_j) + O_p\left(\frac{1}{\sqrt{n}}\right)$, l'hypothèse (H_3^*) entraîne que $\xi_2(n)$ tend vers 0 quand n croît.

Le premier terme $\xi_1(n)$ se décompose en la somme d'un terme $\xi'_1(n)$ relatif à $[AB]$ et d'un terme $\xi''_1(n)$ relatif au complémentaire de $[AB]$ dans $]01[$.

Les $X_{(i)}$ correspondant à des valeurs de i telles que $\frac{i}{n+1}$ soit compris entre A et B sont majorés, en valeur absolue, par $\max(|F^{-1}(A)|, |F^{-1}(B)|) + O_p\left(\frac{1}{\sqrt{n}}\right)$.
Donc $|\xi'_1(n)| < nL \circ \left(\frac{1}{n}\right) + o_p\left(\frac{1}{\sqrt{n}}\right)$ d'après l'hypothèse (H_2^*) , pour une constante L . Donc $\xi'_1(n)$ tend vers 0 quand n croît.

D'autre part $\xi''_1(n)$ se compose de deux termes analogues, l'un relatif à $]0A[$ et l'autre relatif à $]B1[$. Considérons le premier : son comportement asymptotique peut être obtenu en majorant $|h_n(t) - h(t)|$ par $2 \frac{C'}{t^\beta (1-t)^\beta}$ et en considérant que $X_{(i)} = F^{-1}\left(\frac{i}{n+1}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)$. On peut donc majorer $|\xi''_1(n)|$ par $L\sqrt{n} \int_0^A \frac{1}{t^{\beta+b}} dt$ pour une constante L . Comme $b + \beta < \frac{1}{2}$, il en résulte que $\xi''_1(n) \xrightarrow[n \rightarrow \infty]{} 0$.

4 - ESTIMATEUR ASYMPTOTIQUEMENT EFFICACE.

Soit X_1, \dots, X_n un n -échantillon d'une loi sur \mathbb{R} de fonction de répartition $F_\theta = F(\cdot - \theta)$, où F est connue, symétrique ($F(-x) = 1 - F(x)$, $x \in \mathbb{R}$), admet une densité f et vérifie l'hypothèse (H^1) :

(H^1_1) f est trois fois continûment différentiable et strictement positive sur son support.

(H^1_2) l'information de f , $I(f) = \int_{-\infty}^{+\infty} \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx$ est finie

(H^1_3) $(-\log f)'' \circ F^{-1}(t)$ est à variation bornée sur les compacts de $]0,1[$ et il existe un compact $[AB]$ de $]0,1[$ en dehors duquel $(-\log f)'' \circ F^{-1}(t)$ est soit nul, soit dérivable et de dérivée majorée, en valeur absolue par $\frac{C'}{t^{1+\beta}(1-t)^{1+\beta}}$ où C' et β sont deux constantes réelles.

(H^1_4) Il existe deux constantes, C positive quelconque et b telle que $b+\beta < \frac{1}{2}$, telles que, hors du compact $[AB]$, on ait

$$\left| \frac{1}{f(F^{-1}(t))} \right| < \frac{C}{t^{1+b}(1-t)^{1+b}}.$$

Toutes ces hypothèses sur F sont destinées à assurer la réalisation de l'hypothèse (H) pour le couple $(F, h = \frac{1}{I(f)} (-\log f)'' \circ F^{-1})$ car nous allons voir que $h = \frac{1}{I(f)} (-\log f)'' \circ F^{-1}$ est la fonction de poids optimale correspondant à F :

Théorème 4.1 :

Si (X_1, X_2, \dots, X_n) est un n -échantillon d'une loi sur \mathbb{R} de fonction de répartition $F(\cdot - \theta)$ où F est symétrique, connue et vérifie l'hypothèse (H^1) , la suite de statistiques

$$\hat{\theta}_n(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n h\left(\frac{i}{n+1}\right) X_{(i)}$$

où $h(t) = \frac{1}{I(f)} (-\log f)'' \circ F^{-1}(t)$ $t \in]0,1[$

est une suite d'estimateurs de θ , asymptotiquement efficace.

Démonstration :

On cherche, parmi les estimateurs sans biais de θ qui sont de la forme

$$T_n = \frac{1}{n} \sum_{i=1}^n h\left(\frac{i}{n+1}\right) X_{(i)} \text{ un estimateur qui soit de variance minimale.}$$

Il faut donc trouver parmi les fonctions h définies sur $]0,1[$ qui vérifient l'hypothèse (H) et qui réalisent, pour tout θ réel, $\int_0^1 h(t) F_\theta^{-1}(t) dt = \theta$, une fonction qui minimise $\sigma^2 = \int_0^1 K^2(t) dt - \left(\int_0^1 K(t) dt\right)^2$. Rappelons que $dK(t) = h(t) dF^{-1}(t)$.

On cherchera h parmi les fonctions définies sur $]0,1[$ qui vérifient les conditions suivantes :

$$(i) \quad \int_0^1 h(t) dt = 1$$

$$(ii) \quad \int_0^1 K(t) dt = 0$$

$$(iii) \quad \int_0^1 K^2(t) dt \text{ minimum.}$$

En remarquant que $h(t) = \frac{dK(t)}{dF^{-1}(t)}$ et que $dF^{-1}(t) = \frac{1}{f(F^{-1}(t))} dt$, on peut écrire

(i) sous la forme

$$\begin{aligned} (i') \quad \int_0^1 f[F^{-1}(t)] dK(t) &= [f[F^{-1}(t)] K(t)]_0^1 + \int_0^1 K(t) \frac{f'(F^{-1}(t))}{f(F^{-1}(t))} dt \\ &= \int_0^1 K(t) \frac{f'(t)}{f(F^{-1}(t))} dt = 1 \end{aligned}$$

L'inégalité de SCHWARTZ donne un minorant de $\int_0^1 K^2(t) dt$:

$$\int_0^1 K^2(t) dt \int_0^1 \left(\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}\right)^2 dt \geq \left(\int_0^1 K(t) \frac{f'(F^{-1}(t))}{f(F^{-1}(t))} dt\right)^2 = 1$$

Donc

$$\int_0^1 K^2(t) dt \geq \frac{1}{\int_0^1 \left(\frac{f'(F^{-1}(t))}{f(F^{-1}(t))}\right)^2 dt} = \frac{1}{I(f)}$$

ce qui prouve que $\sigma^2 = \int_0^1 K^2(t) dt$ ne peut atteindre son minimum $\frac{1}{I(f)}$ que si

$$K(t) = L \frac{f'(F^{-1}(t))}{f(F^{-1}(t))}, \text{ où } L \text{ est une constante réelle.}$$

Il reste à vérifier les conditions (i)' et (ii) :

$$(ii) \quad L \int_0^1 \left(- \frac{f'(F^{-1}(t))}{f(F^{-1}(t))} \right) dt = - L \int_{-\infty}^{+\infty} f'(x) dx = 0$$

$$(i)' \quad - \int_0^1 K(t) \frac{f'(F^{-1}(t))}{f(F^{-1}(t))} dt = \frac{1}{L} \int_0^1 K^2(t) dt = \frac{1}{2 I(f)} = 1$$

Donc $L = \frac{1}{I(f)}$ et la fonction cherchée vaut $h(t) = \frac{1}{I(f)} (-\text{Log } f)'' \circ F^{-1}(t)$.

Comme il est facile de voir que le couple (F, h) où $h = \frac{1}{I(f)} (-\text{Log } f)'' \circ F^{-1}$ vérifie l'hypothèse (H) lorsque F vérifie (H'), le théorème en résulte.

Exemple :

Si F est la fonction de répartition de la loi gaussienne centrée $N(0, \sigma^2)$ sur \mathbb{R} , la fonction $h(t) = \frac{1}{I(f)} (-\text{Log } f)'' \circ F^{-1}(t) = 1$ pour tout t de $[0, 1]$, et on retrouve la moyenne empirique $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_{(i)}$ comme estimateur asymptotiquement efficace de θ . L'hypothèse (H') est clairement vérifiée par F .

Catherine HUBER

Département Mathématiques
Université Paris-Nord
Avenue J.B. Clément
93430 VILLETANEUSE

CHAPITRE VII
TESTS MINIMAX ROBUSTES

par

Lucien BIRGÉ.

0 - INTRODUCTION.

On veut étudier les tests de deux hypothèses H_0 et H_1 , l'hypothèse H_i ($i = 0, 1$) étant que la loi des variables aléatoires X_n , que l'on observe, est P_i . Un problème se pose lorsque la loi réelle des v. a. X_n n'est pas exactement P_0 ou P_1 , mais une loi voisine, en raison par exemple des erreurs de mesure. Dans ce cas on est amené à rechercher des tests "robustes", c'est-à-dire peu sensibles à ces variations.

L'exposé ci-après est fondé sur les travaux de P.J. HUBER, [45], [46], [51], qui adopte le point de vue du minimax consistant à optimiser les pires performances des procédures de test. On reprend pour l'essentiel l'article de P.J. HUBER et V. STRASSEN [51], en ajoutant une application à certains voisinages classiques d'une probabilité, sur un espace non compact, ce qui permet d'étendre le résultat principal à des voisinages de PROHOROV sur \mathbb{R}^n par exemple. On montre aussi que le problème se ramène à la recherche d'expériences "minimales" au sens de la théorie de BLACKWELL et LE CAM. Un contre-exemple prouve que le résultat obtenu ne peut s'étendre à des problèmes de décision comportant plus de deux hypothèses.

Soient un espace mesuré (E, \mathcal{A}) et \mathcal{M} l'ensemble des mesures de probabilité sur E . Pour tester les hypothèses H_0 et H_1 , on observe des v.a. X_1, \dots, X_n d'un espace Ω dans E .

L'hypothèse H_i ($i = 0, 1$) est que les lois de ces variables appartiennent toutes à un même sous-ensemble \mathcal{P}_i de \mathcal{M} . On suppose que \mathcal{P}_0 et \mathcal{P}_1 sont disjoints. Le plus souvent, ce sont des voisinages, pour une certaine topologie sur \mathcal{M} , de lois de probabilité P_0 et P_1 respectivement.

On notera $\mathcal{P}_i^{\otimes n}$ ($i = 0, 1$) l'ensemble des lois P des variables (X_1, \dots, X_n) , c'est-à-dire :

$$\mathcal{P}_i^{\otimes n} = \{ P = T_1 \otimes T_2 \dots \otimes T_n \mid T_j \in \mathcal{P}_i ; j = 1, \dots, n \} \quad i = 0, 1$$

DÉFINITIONS (0.1) :

Un test randomisé d'ordre n de H_0 contre H_1 est une fonction $\varphi : E^n \rightarrow [0, 1]$; $\varphi(x_1, \dots, x_n)$ représente la probabilité de rejeter H_0 si on a observé (x_1, \dots, x_n) .

Le risque obtenu pour la procédure φ lorsque la loi de (X_1, \dots, X_n) est P s'écrit :

$$R(P, \varphi) = \begin{cases} E_P(\varphi) & \text{si } P \in \mathcal{P}_0^{\otimes n} \\ E_P(1 \dots \varphi) & \text{si } P \in \mathcal{P}_1^{\otimes n} \end{cases}$$

Le niveau α du test, et sa puissance β s'écrivent :

$$\alpha = \sup_{P \in \mathcal{P}_0^{\otimes n}} R(P, \varphi)$$

$$\beta = 1 - \sup_{P \in \mathcal{P}_1^{\otimes n}} R(P, \varphi)$$

On recherche des procédures φ optimales dans chacun des cas suivants :

$$\text{a) Minimiser } \sup_{i=0, 1} \sup_{P_i \in \vartheta_i^{\otimes n}} R(P_i, \varphi)$$

$$\text{b) Minimiser } \sup_{P_1 \in \vartheta_1^{\otimes n}} R(P_1, \varphi) \text{ sous la contrainte}$$

$$\sup_{P_0 \in \vartheta_0^{\otimes n}} R(P_0, \varphi) \leq \alpha \quad \alpha \in [0, 1]$$

$$\text{c) Minimiser } \sup_{P_0 \in \vartheta_0^{\otimes n}} \lambda R(P_0, \varphi) + (1 - \lambda) R(P_1, \varphi) \quad \lambda \in [0, 1]$$

$$P_1 \in \vartheta_1^{\otimes n}$$

Ces trois problèmes ont les mêmes familles de solutions, (a) et (c) se ramenant à (b) pour un choix convenable de α . Les solutions φ du problème (b) relatives au nombre α seront dites "tests minimax de niveau α ".

On sait résoudre ce problème si ϑ_0 et ϑ_1 sont réduits chacun à un seul élément P_0 et P_1 respectivement. D'après le lemme de NEYMAN-PEARSON [68], la solution est alors un test de la forme suivante :

$$\varphi(x_1, \dots, x_n) = 1 \quad \text{si} \quad \prod_{j=1}^n \frac{dP_1}{dP_0}(x_j) > C$$

$$\varphi(x_1, \dots, x_n) = \lambda \quad \text{si} \quad \prod_{j=1}^n \frac{dP_1}{dP_0}(x_j) = C$$

$$\varphi(x_1, \dots, x_n) = 0 \quad \text{si} \quad \prod_{j=1}^n \frac{dP_1}{dP_0}(x_j) < C$$

avec $C \geq 0$, $\lambda \in [0, 1]$ tels que $E_{P_0^{\otimes n}}(\varphi) = \alpha$. Un tel test est dit "test de rapport de vraisemblance de P_0 contre P_1 ".

Si l'on peut trouver un couple (Q_0, Q_1) dans $\vartheta_0 \times \vartheta_1$ tel que pour tout test φ de rapport de vraisemblance de Q_0 contre Q_1 , $R(Q_i^{\otimes n}, \varphi)$ maximise $R(P_i, \varphi)$ pour P_i dans $\vartheta_i^{\otimes n}$ ($i = 0, 1$), les tests de rapport de vraisemblance de Q_0 contre Q_1 seront des tests minimax de H_0 contre H_1 .

DÉFINITION (0.2) :

Un couple $(Q_0, Q_1) \in \mathcal{Q}_0 \times \mathcal{Q}_1$ sera dit "minimal" s'il vérifie les conditions suivantes:

Pour tout test φ de rapport de vraisemblance entre Q_0 et Q_1 on a :

$$\forall P \in \mathcal{P}_i^{\otimes n} \quad R(P, \varphi) \leq R(Q_i^{\otimes n}, \varphi) \quad i = 0, 1$$

PROPOSITION (0.3) :

Si (Q_0, Q_1) est un couple minimal, les tests de rapport de vraisemblance de Q_0 contre Q_1 sont des tests minimax de H_0 contre H_1 .

Remarque : Les définitions de $\mathcal{Q}_0^{\otimes n}$ et $\mathcal{Q}_1^{\otimes n}$ autorisent les variables indépendantes X_i à avoir des lois distinctes.

1 - DISTRIBUTIONS MINIMALES.

Donnons une condition suffisante pour que le couple (Q_0, Q_1) soit minimal.

PROPOSITION (1.1) :

Soit γ une version du rapport de vraisemblance $\frac{dQ_1}{dQ_0}$. Si pour tout $P_i \in \mathcal{P}_i$ ($i = 0, 1$) et tout réel t , on a les inégalités

$$(1) \quad P_0 [\gamma > t] \leq Q_0 [\gamma > t] \leq Q_1 [\gamma > t] \leq P_1 [\gamma > t]$$

le couple (Q_0, Q_1) est minimal.

Preuve : Ceci est une conséquence du théorème suivant.

THÉORÈME (1.2) :

Si le couple (Q_0, Q_1) vérifie les inégalités (1), pour tout entier n et tout réel $\alpha \in [0, 1]$, le test de rapport de vraisemblance de niveau α entre Q_0 et Q_1 défini par les équations :

$$\begin{aligned} \varphi(x_1, \dots, x_n) &= 1 & \text{si} & \quad \prod_{j=1}^n \gamma(x_j) > C \\ \varphi(x_1, \dots, x_n) &= \lambda & \text{si} & \quad \prod_{j=1}^n \gamma(x_j) = C \\ \varphi(x_1, \dots, x_n) &= 0 & \text{si} & \quad \prod_{j=1}^n \gamma(x_j) < C \end{aligned}$$

où λ et C sont choisis de manière que $E_{Q_0 \otimes n}(\varphi) = \alpha$ est un test minimax de H_0 contre H_1 de niveau α , ayant la même puissance que le test de Q_0 contre Q_1 .

Preuve : L'existence de λ et C est assurée par le lemme de NEYMAN-PEARSON [68].

On va se contenter de montrer que l'on a :

$$\forall P \in \mathcal{P}_0^{\otimes n} \quad R(P, \varphi) \leq R(Q_0^{\otimes n}, \varphi) = \alpha$$

Un calcul analogue prouverait que l'on a aussi :

$$\forall P \in \mathcal{P}_1^{\otimes n} \quad R(P, \varphi) \leq R(Q_1^{\otimes n}, \varphi)$$

Si P est un élément de $\mathcal{P}_0^{\otimes n}$, il s'écrit

$$P = T_1 \otimes \dots \otimes T_n \quad T_j \in \mathcal{P}_0 \quad j = 1, \dots, n$$

et d'après (1) $F_j(t) = T_j\{\gamma > t\} \leq Q_0\{\gamma > t\} = G(t) \quad j = 1, \dots, n$

$$R(P, \varphi) = \mathbb{E}_P(\varphi) = P\left\{x \in E^n \mid \prod_{j=1}^n \gamma(x_j) > C\right\} + \lambda P\left\{x \in E^n \mid \prod_{j=1}^n \gamma(x_j) = C\right\}$$

D'après [68], il existe des fonctions croissantes au sens large g, f_j ($j = 1, \dots, n$), une variable aléatoire V de \mathbb{R} dans \mathbb{R} et une probabilité S sur \mathbb{R} telles que l'on ait :

$$f_j(v) \leq g(v) \quad j = 1, \dots, n$$

$$S\{f_j(V) > t\} = F_j(t) \quad S\{g(V) > t\} = G(t)$$

Soient V_j ($j = 1, \dots, n$) des répliques indépendantes équidistribuées de V . On peut alors écrire :

$$\begin{aligned} R(P, \varphi) &= S^{\otimes n} \left[\prod_{j=1}^n f_j(V_j) > C \right] + \lambda S^{\otimes n} \left[\prod_{j=1}^n f_j(V_j) = C \right] \\ &= \lambda S^{\otimes n} \left[\prod_{j=1}^n f_j(V_j) \geq C \right] + (1 - \lambda) S^{\otimes n} \left[\prod_{j=1}^n f_j(V_j) > C \right] \\ &\leq \lambda S^{\otimes n} \left[\prod_{j=1}^n g(V_j) \geq C \right] + (1 - \lambda) S^{\otimes n} \left[\prod_{j=1}^n g(V_j) > C \right] \\ &= \mathbb{E}_{Q_0^{\otimes n}}(\varphi) = R(Q_0^{\otimes n}, \varphi) = \alpha \end{aligned} \quad \text{C.Q.F.D.}$$

On observe une suite de v. a. $X_j : \Omega \rightarrow E$, indépendantes, dont les lois appartiennent toutes à \mathcal{P}_0 (hypothèse H_0) ou à \mathcal{P}_1 (hypothèse H_1). Un test séquentiel φ de H_0 contre H_1 est une fonction φ de $E^{\mathbb{N}} \rightarrow [0, 1]$ associée à un temps d'arrêt $N(\omega)$ de $\Omega \rightarrow \mathbb{N}$. C'est-à-dire qu'il existe une fonction M de $E^{\mathbb{N}} \rightarrow \mathbb{N}$ telle que l'on ait :

$$N(\omega) = M(X_1(\omega), \dots, X_n(\omega), \dots)$$

et pour tout n l'ensemble $\{x \in \mathbb{E}^N \mid M(x) \leq n\}$ ne dépend que des coordonnées (x_1, \dots, x_n) de x . φ est alors la probabilité de rejeter H_0 lorsque l'on a observé $(x_1, x_2, \dots, x_M(x))$. On définit le risque comme précédemment :

$$R(P, \varphi) = \begin{cases} \mathbb{E}_P(\varphi) & \text{si } P \in \varphi_0^{\otimes n} \\ \mathbb{E}_P(1 - \varphi) & \text{si } P \in \varphi_1^{\otimes n} \end{cases}$$

D'après [68] si φ_0 et φ_1 sont chacun réduits à un élément, les tests optimaux sont des tests de rapport de vraisemblance de la forme suivante :

$$M(x) = \inf \left\{ n \mid \prod_{j=1}^n \gamma(x_j) \leq K' \text{ ou } \prod_{j=1}^n \gamma(x_j) \geq K'' \right\} \quad 0 \leq K' < K''$$

On choisit H_0 ($\varphi = 0$) si $\prod_{j=1}^N \gamma(X_j(\omega)) \leq K'$, H_1 ($\varphi = 1$) dans l'autre cas, K' et K'' étant tels que $\mathbb{E}_{P_i}(M)$ soit finie ($i = 0, 1$).

THEOREME (1.3) :

a) Si le couple (Q_0, Q_1) vérifie (1), les tests séquentiels de rapport de vraisemblance, de niveau α et puissance β , de Q_0 contre Q_1 sont des tests minimax de H_0 contre H_1 au niveau α et de même puissance β .

b) Si le logarithme du rapport de vraisemblance $\gamma = \frac{dQ_1}{dQ_0}$ est borné et si on a $\mathbb{E}_P(M) < +\infty \quad \forall P \in \varphi_0^{\otimes n} \cup \varphi_1^{\otimes n}$ alors

$$\lim_{\substack{K' \rightarrow 0 \\ K'' \rightarrow +\infty}} \frac{\mathbb{E}_{P_i}(M)}{\mathbb{E}_{Q_i}(M)} \leq 1 \quad \forall P_i \in \varphi_i^{\otimes N} \quad i = 0, 1$$

avec $\overline{Q_i} = \bigotimes_{j=0}^{\infty} Q_i$, la limite étant prise en faisant tendre K' vers 0 et K'' vers l'infini de manière que le rapport $\left| \frac{\log K'}{\log K''} \right|$ reste compris entre deux nombres strictement positifs.

Preuve :

a) Soit φ un test séquentiel de rapport de vraisemblance de Q_0 contre Q_1 défini

par $K' < K''$. On a en reprenant les notations du théorème (1.2) et en posant :

$$\begin{aligned}
 P &= \bigotimes_{j=1}^{\infty} T_j & T_j &\in \vartheta_0 & \forall j & & \bar{S} &= \bigotimes_{j=1}^{\infty} S \\
 \mathbb{E}_P(\varphi) &= \mathbb{E}_P \left[x \in E^{\mathbb{N}} \mid \prod_{j=1}^{M(x)} \gamma(x_j) \geq K'' \right] \\
 &= \mathbb{E}_P \left[x \in E^{\mathbb{N}} \mid \exists p, \prod_{j=1}^p \gamma(x_j) \geq K'' \text{ et } \prod_{j=1}^i \gamma(x_j) > K' \quad \forall i < p \right] \\
 &= \bar{S} \left\{ \exists p, \prod_{j=1}^p f_j(V_j) \geq K'' \quad \text{et} \quad \prod_{j=1}^i f_j(V_j) > K' \quad \forall i < p \right\} \\
 &\leq \bar{S} \left\{ \exists p, \prod_{j=1}^p g(V_j) \geq K'' \quad \text{et} \quad \prod_{j=1}^i g(V_j) > K' \quad \forall i < p \right\} \\
 &= \mathbb{E}_{\bar{Q}_0} \left[x \in E^{\mathbb{N}} \mid \prod_{j=1}^{M(x)} \gamma(x_j) \geq K'' \right]
 \end{aligned}$$

Donc $R(P, \varphi) \leq R(\bar{Q}_0, \varphi)$ et de même $R(P, \varphi) \leq R(\bar{Q}_1, \varphi)$ si $P \in \vartheta_1^{\otimes n}$.

b) Par hypothèse $|\text{Log } \gamma| < A$. On pose :

$$\text{Log } \gamma(X_j) = Z_j \quad \sum_{j=1}^p Z_j = Y_p = \text{Log} \left[\prod_{j=1}^p \gamma(X_j) \right]$$

On va montrer que
$$\lim_{\substack{K' \rightarrow 0 \\ K'' \rightarrow \infty}} \frac{\mathbb{E}_P[M]}{\mathbb{E}_{\bar{Q}_0}[M]} \leq 1 \quad \text{si } P \in \vartheta_0^{\otimes n}$$

Avec les hypothèses $|Z_j| < A$ et $\mathbb{E}_P[M] < +\infty$, on a d'après un résultat classique [68] utilisant la formule de FUBINI

$$\mathbb{E}_P[Y_M] = \sum_{j=1}^{\infty} \mathbb{E}_P[Z_j \cdot 1_{(M \geq j)}] = \sum_{j=1}^{\infty} \mathbb{E}_{T_j}[Z_j] P(M \geq j)$$

puisque $1_{(M \geq j)}$ est indépendant de Z_j , et aussi

$$\mathbb{E}_{\bar{Q}_0}[Y_M] = \mathbb{E}_{Q_0}[Z_1] \sum_{j=1}^{\infty} Q_0(M \geq j) = \mathbb{E}_{Q_0}[Z_1] \mathbb{E}_{\bar{Q}_0}[M]$$

D'après (1)
$$\mathbb{E}_{T_j}[Z_j] \leq \mathbb{E}_{Q_0}[Z_j] \quad \forall j$$

et
$$\mathbb{E}_{Q_0}[Z_j] = \mathbb{E}_{Q_0}[\text{Log } \gamma(X_j)] < \text{Log } \mathbb{E}_{Q_0}[\gamma(X_j)] = \text{Log} \int \frac{dQ_1}{dQ_0} dQ_0 \leq 0$$

$$\text{Donc } \mathbb{E}_P[Y_M] \leq \sum_{j=1}^{\infty} \mathbb{E}_{Q_0}[Z_j] P(M \geq j) = \mathbb{E}_{Q_0}[Z_1] \mathbb{E}_P[M] \leq 0$$

$$\text{et } \frac{\mathbb{E}_P[M]}{\mathbb{E}_{Q_0}[M]} = \frac{\mathbb{E}_{Q_0}[Z_1] \mathbb{E}_P[M]}{\mathbb{E}_{Q_0}[Z_1] \mathbb{E}_{Q_0}[M]} \leq \frac{\mathbb{E}_P[Y_M]}{\mathbb{E}_{Q_0}[Y_M]} \text{ car } \mathbb{E}_{Q_0}[Y_M] < 0$$

$$\mathbb{E}_P[Y_M] = \mathbb{E}_P[Y_M \mathbf{1}_{(Y_M \leq \text{Log } K')}] + \mathbb{E}_P[Y_M \mathbf{1}_{(Y_M \geq \text{Log } K'')}]$$

$$\text{sur } Y_M \leq \text{Log } K' \quad Y_M \underset{K' \rightarrow 0}{\sim} \text{Log } K'$$

$$\text{sur } Y_M \geq \text{Log } K'' \quad Y_M \underset{K'' \rightarrow \infty}{\sim} \text{Log } K''$$

puisque $|\text{Log } \gamma|$ est borné par A.

$$\overline{Q_0}[Y_M \leq \text{Log } K'] = 1 - \overline{Q_0}[Y_M \geq \text{Log } K''] \xrightarrow{K'' \rightarrow \infty} 1$$

$$\text{d'après l'inégalité (voir [68])} \quad K'' \leq \frac{\overline{Q_1}[Y_M \geq \text{Log } K'']}{\overline{Q_0}[Y_M \geq \text{Log } K'']}$$

La partie (a) implique $P[Y_M \geq \text{Log } K''] \leq \overline{Q_0}[Y_M \geq \text{Log } K'']$.

$$\text{Donc } P[Y_M \geq \text{Log } K''] \xrightarrow{K'' \rightarrow \infty} 0 \text{ et } P[Y_M \leq \text{Log } K'] \xrightarrow{K'' \rightarrow \infty} 1$$

Comme le rapport $|\frac{\text{Log } K'}{\text{Log } K''}|$ reste compris entre deux nombres positifs,

$$\mathbb{E}_P[Y_M] \underset{\substack{K' \rightarrow 0 \\ K'' \rightarrow +\infty}}{\sim} \text{Log } K', \text{ quel que soit } P \text{ dans } \varphi_0^{\otimes N}.$$

$$\text{On a donc } \lim_{\substack{K' \rightarrow 0 \\ K'' \rightarrow +\infty}} \frac{\mathbb{E}_P[M]}{\mathbb{E}_{Q_0}[M]} \leq \lim_{\substack{K' \rightarrow 0 \\ K'' \rightarrow +\infty}} \frac{\mathbb{E}_P[Y_M]}{\mathbb{E}_{Q_0}[Y_M]} = 1$$

On montrerait de façon analogue que l'on a

$$\lim_{\substack{K' \rightarrow 0 \\ K'' \rightarrow +\infty}} \frac{\mathbb{E}_P[M]}{\mathbb{E}_{Q_1}[1]} \leq 1 \quad \forall P \in \varphi_1^{\otimes N}$$

C.Q.F.D.

2 - INTRODUCTION ET PROPRIÉTÉS DES CAPACITÉS.

a) Motivations :

Dans toute la suite E sera un espace séparable, métrique, complet, \mathcal{A} la tribu de ses boréliens, \mathcal{P} l'ensemble des probabilités sur (E, \mathcal{A}) . \mathcal{P}_0 et \mathcal{P}_1 sont deux sous-ensembles disjoints de \mathcal{P} .

Pour étudier les tests de \mathcal{P}_0 contre \mathcal{P}_1 , de région de rejet $A \in \mathcal{A}$, on est amené à calculer

$$\begin{aligned} v_0(A) &= \sup_{\mathcal{P}_0} P(A) && \text{niveau du test} \\ u_1(A) &= \inf_{\mathcal{P}_1} P(A) && \text{puissance du test.} \end{aligned}$$

DÉFINITION (2.1) :

Si \mathcal{C} est une partie de \mathcal{P} , on appellera enveloppe supérieure (respectivement inférieure) de \mathcal{C} la fonction de \mathcal{A} dans $[0, 1]$ ainsi définie :

$$v(A) = \sup_{\mathcal{C}} P(A) \quad (\text{respectivement } u(A) = \inf_{\mathcal{C}} P(A))$$

u et v sont liés par la relation $u(A) + v(A^c) = 1$.

On constate que v_0 et u_1 possèdent des propriétés de monotonie et continuité qui en font des capacités au sens de CHOQUET (voir [18]), pourvu que \mathcal{P}_0 et \mathcal{P}_1 soient suffisamment réguliers. Ceci explique l'introduction des capacités pour résoudre ce problème.

b) Capacités de CHOQUET :

DÉFINITIONS (2.2) :

Soit v une fonction de \mathcal{A} dans $[0, 1]$, on dira que c'est une capacité de CHOQUET si elle vérifie les quatre propriétés :

- (a) $v(\phi) = 0$ $v(E) = 1$
- (b) $A \subset B \quad \Rightarrow \quad v(A) \leq v(B)$
- (c) Pour toute suite croissante d'éléments A_n de \mathcal{A} tendant vers A , $v(A_n)$ tend vers $v(A)$.
- (d) Pour toute suite décroissante de fermés F_n tendant vers F , $v(F_n)$ tend vers $v(F)$.
Si de plus, v vérifie la propriété (e) elle est dite bi-alternative.
- (e) $v(A \cup B) + v(A \cap B) \leq v(A) + v(B) \quad \forall A, B \in \mathcal{A}$.
Si v est une capacité bi-alternative, la fonction u de \mathcal{A} dans $[0, 1]$ définie par $u(A) = 1 - v(A^c)$ est dite capacité bi-monotone conjuguée de v .

Une capacité bi-monotone n'est pas une capacité au sens de (a - b - c - d) mais elle possède les propriétés (a) (b) ainsi que les suivantes :

- (c') Pour toute suite décroissante d'éléments A_n de \mathcal{A} tendant vers A , $u(A_n)$ tend vers $u(A)$.
- (d') Pour toute suite croissante d'ouverts G_n tendant vers G , $u(G_n)$ tend vers $u(G)$.
- (e') $u(A \cap B) + u(A \cup B) \geq u(A) + u(B) \quad \forall A, B \in \mathcal{A}$.

Remarques :

- v est une probabilité si et seulement si on a égalité dans (e) pour tous les ensembles A, B .
- l'inégalité (e) appliquée à $B = A^c$ montre que $u(A) \leq v(A)$.
- on voit immédiatement que les enveloppes supérieures de parties de \mathcal{A} vérifient (a - b - c).

DÉFINITION (2.3) : voir [14] - [18] - [19] - [23] - [71]

Un élément A de \mathcal{A} est dit v -capacitable s'il vérifie :

$$v(A) = \sup_{K \subset A} v(K) = \inf_{G \supset A} v(G) \quad K \text{ compact, } G \text{ ouvert}$$

THÉORÈME (2.4) : [14] - [18]

Toute partie borélienne de E est v-capacitable pour toute capacité v sur \mathcal{A} .

c) Fonctionnelles associées :

Soit ψ une fonction monotone, positive, bornée sur \mathcal{A} telle que $\psi(\emptyset) = 0$. Pour toute fonction f positive, mesurable, bornée sur E on définit l'intégrale de f par

$$\tilde{\psi}(f) = \int_0^\infty \psi\{x \mid f(x) > t\} dt = \int_0^\infty \psi\{x \mid f(x) \geq t\} dt$$

$\tilde{\psi}$ est une fonctionnelle monotone, positive et positivement homogène. Si \mathcal{C}^+ désigne l'espace des fonctions continues, positives, bornées sur E, muni de la topologie de la convergence uniforme, $\tilde{\psi}$ est continue sur \mathcal{C}^+ .

D'après [14] - [15] $\tilde{\psi}$ est sous-additive ($\tilde{\psi}(f+g) \leq \tilde{\psi}(f) + \tilde{\psi}(g)$) si et seulement si ψ est bi-alternative (propriété (e)). En particulier, si ψ est une mesure, $\tilde{\psi}$ est additive et l'on a $\tilde{\psi}(f) = \int f d\psi$ d'après le théorème de FUBINI.

LEMME (2.5) :

Si ψ possède la propriété (d) de continuité sur les suites décroissantes de fermés et si tout borélien de E vérifie la propriété de capacabilité :

$$\psi(A) = \sup_{K \subset A} \psi(K) \quad K \text{ compact}$$

les valeurs de ψ sur \mathcal{A} sont déterminées dès que l'on connaît celles de $\tilde{\psi}$ sur \mathcal{C}^+ .

Si de plus φ est une autre fonctionnelle ayant ces propriétés :

$$\tilde{\varphi} \leq \tilde{\psi} \text{ entraîne } \varphi(A) \leq \psi(A) \quad \forall A \in \mathcal{A}.$$

Preuve :

ψ est déterminée par ses valeurs sur les compacts et vérifie

$$\psi(K) = \int_0^\infty \psi\{x \mid 1_K \geq t\} dt = \tilde{\psi}(1_K)$$

Soit alors (f_n) une suite de fonctions de \mathcal{C}^+ décroissant vers 1_K . Par exemple :

$$\begin{aligned} f_n(x) &= 1 - n d(x, K) & \text{si } d(x, K) \leq \frac{1}{n} \\ &= 0 & \text{si } d(x, K) > \frac{1}{n} \end{aligned}$$

où d est une distance sur E .

Les fermés $F_n = \{f_n \geq t\}$ décroissent vers $\{1_K \geq t\}$. Donc $\psi(F_n)$ tend vers $\psi(\{1_K \geq t\})$ et on a par convergence monotone

$$\lim_n \tilde{\psi}(f_n) = \lim_n \int_0^\infty \psi\{f_n \geq t\} dt = \int_0^\infty \psi\{1_K \geq t\} dt = \tilde{\psi}(K)$$

La deuxième assertion s'en déduit immédiatement.

C.Q.F.D.

3 - LIEN ENTRE LES CAPACITÉS ET LES PROBABILITÉS QU'ELLES MAJORENT.

Soit ν une capacité sur l'espace (E, \mathcal{A}) (séparable, métrisable, complet). On note $\mathcal{C}_\nu = \{P \in \mathcal{M} \mid P \leq \nu\}$ et $\sup \mathcal{C}$ l'enveloppe supérieure de $\mathcal{C} \subset \mathcal{M}$. On veut étudier les rapports qui existent entre ν et \mathcal{C}_ν , \mathcal{C} et $\sup \mathcal{C}$.

LEMME (3.1) :

Si ν est une capacité, l'égalité $\tilde{P}(f) = \int f dP$ définit une bijection entre les probabilités P majorées par ν et les fonctionnelles linéaires positives \tilde{P} sur \mathcal{C} qui vérifient $\tilde{P}(1) = 1$ et $\tilde{P}(f) \leq \tilde{\nu}(f) \quad \forall f \in \mathcal{C}^+$.

Preuve :

Dans un sens le résultat est évident. Il suffit donc de montrer que \tilde{P} induit une probabilité qui sera automatiquement majorée par ν d'après le lemme (2.5). D'après un théorème de DANIELL [73], il suffit de vérifier que si la suite (f_n) tend vers 0 en décroissant, $\tilde{P}(f_n)$ tend vers 0.

Comme $\{f_n \geq t\}$ est fermé et tend vers \emptyset , $\nu\{f_n \geq t\}$ tend vers 0 d'après (d) et par convergence monotone :

$$\tilde{P}(f_n) \leq \tilde{\nu}(f_n) = \int_0^\infty \nu\{f_n \geq t\} dt \rightarrow 0$$

C.Q.F.D.

LEMME (3.2) :

Si ν est une capacité, l'ensemble $\mathcal{C}_\nu = \{P \in \mathcal{M} \mid P \leq \nu\}$ est convexe et faiblement compact.

Preuve :

On dit qu'un sous-ensemble \mathcal{C} de \mathcal{M} est tendu si pour tout ϵ positif on peut trouver un compact K tel que

$$P(K^c) < \epsilon \quad \forall P \in \mathcal{C}.$$

D'après [9], si φ_v est faiblement fermé et tendu il est faiblement compact.

- Montrons que φ_v est tendu.

Soit d une distance sur E . D'après [9], E est tendu si et seulement si pour tout ϵ , on peut trouver $\delta > 0$ et S réunion finie de sphères de rayon δ tels que $P(S^C) \leq \epsilon$, $\forall P \in \varphi_v$.

Si (x_n) est une suite dense dans E on définit la suite (g_n) de \mathcal{C}^+ de façon que

$$g_n(x) = 0 \quad \text{si } d(x_n, x) \leq \frac{\delta}{2} \quad \quad g_n(x) = 1 \quad \text{si } d(x_n, x) \geq \delta$$

$$(\text{par exemple } g_n(x) = \frac{2}{\delta} [d(x_n, x) - \frac{\delta}{2}] \quad \text{pour } \frac{\delta}{2} \leq d(x_n, x) \leq \delta)$$

$$\text{On pose} \quad f_n = \inf \{g_m \mid m \leq n\} \quad \quad S_n = \{x \mid f_n(x) < 1\}$$

$$S_n \subset \{x \mid \exists x_m, m \leq n \text{ et } d(x_m, x) < \delta\}$$

S_n est donc inclus dans une réunion finie de sphères de rayon δ .

$$P(S_n^C) \leq v(S_n^C) = v\{f_n \geq 1\} \leq \int_0^\infty v\{f_n \geq t\} dt = \tilde{v}(f_n)$$

Comme (x_n) est une suite dense, f_n tend vers 0 en décroissant et donc $\tilde{v}(f_n)$ tend vers 0.

- Par la bijection du Lemme (3.1), la convergence faible de P équivaut à la convergence simple de \tilde{P} sur \mathcal{C}^+ et φ_v est en bijection avec $\{\tilde{P} \mid \tilde{P} \leq \tilde{v}\}$ qui est fermé pour la convergence simple, il est donc faiblement fermé.

-La convexité est claire.

LEMME (3.3) :

Si $v = \sup \varphi$ et si φ est compact, alors (d) est vérifiée et v est une capacité.

Preuve :

Si (d) n'était pas vraie, on pourrait trouver une suite F_n de fermés décroissant vers F et un nombre α tels que l'on ait :

$$\forall n \quad v(F_n) > \alpha > v(F)$$

et donc on peut trouver des éléments P_n de \mathcal{P} tels que

$$\forall n \quad P_n(F_n) > \alpha$$

Si P est un point de \mathcal{P} adhérent à la suite (P_n) et (P_{n_k}) une sous-suite tendant vers P , on a :

$$P(F_n) \geq \overline{\lim}_k P_{n_k}(F_n) \quad \text{car } F_n \text{ est fermé (voir [9])}$$

$$\text{donc} \quad P(F_n) \geq \overline{\lim}_k P_{n_k}(F_{n_k}) \geq \alpha \quad \forall n$$

On aurait alors $P(F) \geq \alpha$ et $v(F) < \alpha$, ce qui est contradictoire.

PROPOSITION (3.4) :

Si v est une capacité bi-alternative et h une fonction semi-continue supérieurement (s.c.s.) sur E , il existe une probabilité Q dans \mathcal{P}_v telle que pour tout t on ait $Q\{h > t\} = v\{h > t\}$ et $Q\{h \geq t\} = v\{h \geq t\}$. En particulier \mathcal{P}_v est non vide.

Preuve :

On peut toujours supposer que h est > 0 et bornée, quitte à la remplacer par $h' = 2 + \text{Arc tg } h$, les familles d'ensembles $\{h > t\}$, $t \in \mathbb{R}$ et $\{h' > t'\}$, $t' \in \mathbb{R}$ étant identiques.

Si h est s.c.s., elle est limite décroissante d'une suite de fonctions g_n de \mathcal{C}^+ [13]. Comme les ensembles $\{g_n \geq t\}$ sont fermés, $\tilde{v}(g_n)$ tend vers $\tilde{v}(h)$. On considère alors les ensembles

$$U = \{f \in \mathcal{C} \mid \tilde{v}(|f|) < \tilde{v}(h)\} \quad V = \{g \in \mathcal{C}^+ \mid g \geq h\}$$

U est un ouvert convexe d'après (e), V est un fermé convexe et leur intersection est vide. D'après le théorème de HAHN-BANACH, il existe une fonctionnelle linéaire continue \tilde{Q} qui vérifie $\tilde{Q}(f) < \tilde{Q}(g)$ pour tout f dans U et tout g dans V . \tilde{Q} étant linéaire prend sur U des valeurs à la fois positives et négatives, donc $\inf_{g \in V} \tilde{Q}(g)$ est

positif, ce qui permet de normaliser \tilde{Q} afin d'avoir :

$$\inf_{g \in V} \tilde{Q}(g) = \inf_{g \in V} \tilde{v}(g) = \tilde{v}(h) \text{ et } \tilde{Q}(f) < \tilde{Q}(g) \quad \forall f \in U, g \in V$$

Si \tilde{Q} n'était pas positive, on pourrait trouver $f \geq 0$ tel que $\tilde{Q}(f)$ soit négatif.

Il existerait alors $\lambda > 0$, tel que

$$\tilde{Q}(g + \lambda f) = \tilde{Q}(g) + \lambda \tilde{Q}(f) < 0 \text{ et } g + \lambda f \geq h$$

ce qui serait une contradiction.

Si f est dans U , on a $\tilde{v}(|f|) < \tilde{v}(h)$ et $\tilde{Q}(f) \leq \tilde{v}(h)$.

Si $\tilde{Q}(f)$ était supérieur à $\tilde{v}(|f|)$, on pourrait trouver un réel c positif vérifiant : $\tilde{v}(|cf|) < \tilde{v}(h) < \tilde{Q}(cf)$, ce qui entraînerait que cf est dans U et la deuxième inégalité serait contradictoire. On a donc sur U $\tilde{Q}(f) \leq \tilde{v}(|f|)$. Par linéarité cette relation s'étend à \mathcal{C}^+ tout entier.

D'après le théorème de représentation de RIESZ, \tilde{Q} est induite par une mesure $Q \leq v$ d'après le Lemme (2.5). $\tilde{Q}(h) = \tilde{v}(h)$ car on a :

$$\begin{aligned} \tilde{v}(h) &\geq \int h dQ = \lim_n \int g_n dQ = \lim_n \tilde{Q}(g_n) \geq \tilde{v}(h) \quad (g_n \in V) \\ \int_0^\infty Q(h > t) dt &= \int_0^\infty v(h > t) dt \end{aligned}$$

On a donc $Q(h > t) = v(h > t)$ p.s. et le résultat est vrai pour tout t par convergence monotone croissante.

$$\text{D'où} \quad Q(E) = Q\{h > 0\} = v(E) = 1.$$

On a aussi :

$$v\{h \geq t\} \leq \lim_{n \rightarrow \infty} v\{h > t - \frac{1}{n}\} = \lim_{n \rightarrow \infty} Q\{h > t - \frac{1}{n}\} = Q\{h \geq t\}$$

ce qui démontre l'égalité $v\{h \geq t\} = Q\{h \geq t\}$

LEMME (3.5) :

Si v est une capacité bi-alternative, pour tout A dans \mathcal{A} , il existe une probabilité dans \mathcal{P}_v telle que $Q(A)$ égale $v(A)$. v est donc l'enveloppe supérieure de \mathcal{P}_v .

Preuve :

Soit une suite croissante (K_n) de compacts inclus dans A tels que $v(A) = \sup_n v(K_n)$. Posons :

$$\begin{aligned} h &= 1 && \text{si } x \in K_1 \\ h &= 1/n && \text{si } x \in K_n - K_{n-1}, \quad \forall n > 1 \\ h &= 0 && \text{si } x \notin \bigcup_n K_n \end{aligned}$$

h est une fonction s.c.s. telle que $K_n = \{h \geq \frac{1}{n}\}$. D'après la proposition (3.4), on peut trouver Q dans \mathcal{P}_v vérifiant $Q(K_n) = v(K_n)$, $\forall n$. On a donc :

$$Q(A) \geq \sup_n Q(K_n) = \sup_n v(K_n) = v(A)$$

C.Q.F.D.

4 - EXEMPLES [51] .

i) Si $v = \sup \vartheta$, ϑ_v peut être strictement plus grand que l'enveloppe convexe fermée de ϑ :

$$E = \{1 ; 2 ; 3\} \quad P_0 = \{\frac{1}{2} ; \frac{1}{2} ; 0\} \quad P_1 = \{\frac{4}{6} ; \frac{1}{6} ; \frac{1}{6}\} \quad v = \sup (P_0, P_1)$$

$$\text{On a } v(1) = \frac{4}{6} \quad v(2) = \frac{1}{2} \quad v(3) = \frac{1}{6} \quad v(1 ; 2) = 1 \quad v(2 ; 3) = \frac{1}{2}$$

$$v(3 ; 1) = \frac{5}{6}$$

v est alors bi-alternative et $\vartheta_v = \{\frac{3+t}{6} ; \frac{3-t-s}{6} ; \frac{1}{6}\}$ avec $0 \leq s, t \leq 1$.

Mais la fermeture convexe de (P_0, P_1) est obtenue pour $s = t$.

ii) v n'est pas nécessairement bi-alternative :

$$E = \{1 ; 2 ; 3 ; 4\} \quad P_0 = \{\frac{5}{10} ; \frac{2}{10} ; \frac{2}{10} ; \frac{1}{10}\} \quad P_1 = \{\frac{6}{10} ; \frac{1}{10} ; \frac{1}{10} ; \frac{2}{10}\}$$

$$\text{si } v = \sup (P_0, P_1) \quad A = (1 ; 2) \quad B = (1 ; 3) \quad \text{on a}$$

$$v(A \cup B) + v(A \cap B) = \frac{9}{10} + \frac{6}{10} = \frac{15}{10} > v(A) + v(B) = \frac{7}{10} + \frac{7}{10} = \frac{14}{10}$$

iii) On suppose ici que l'espace E est compact et que $P_0 \in \mathcal{M} ; \epsilon, \delta > 0$.

On définit les fonctions suivantes :

$$v_1(A) = (1 - \epsilon) P_0(A) + \epsilon \quad \text{si } A \neq \emptyset \quad v_1(\emptyset) = 0$$

$$v_2(A) = \inf (P_0(A) + \epsilon, 1) \quad \text{si } A \neq \emptyset \quad v_2(\emptyset) = 0$$

$$v_3(A) = \inf (P_0(A^\delta) + \epsilon, 1) \quad \text{si } A \neq \emptyset \quad v_3(\emptyset) = 0$$

où A^δ est le voisinage de $A = \{x \mid d(x, A) \leq \delta\}$.

Ces fonctions sont des capacités, bi-alternatives, la propriété (d) étant assurée par la compacité de E .

Les ensembles ϑ_v correspondants sont :

$$\vartheta_{v_1} = \{P = (1 - \epsilon) P_0 + \epsilon H \mid H \in \mathfrak{M}\}$$

$$\vartheta_{v_2} = \{P \in \mathfrak{M} \mid \|P - P_0\| < \epsilon\}$$

$$\vartheta_{v_3} = \{P \in \mathfrak{M} \mid P(A) \leq P_0(A^\delta) + \epsilon \quad \forall A \in \mathcal{A}\}$$

Le premier correspond à un modèle de contamination, les autres à des voisinages de P_0 pour les métriques de la variation totale et de PROHOROV (topologie de la convergence faible) respectivement. Ces trois types d'ensembles sont compacts si E l'est. On trouvera quelques résultats explicites concernant ϑ_{v_1} et ϑ_{v_2} dans le Chapitre VIII du séminaire et [45].

iv) Cas particulier d'un espace E localement compact :

Dans ce cas, la propriété (d) n'est plus vérifiée par les fonctions définies en iii) car il existe des suites décroissantes de fermés non vides tendant vers l'ensemble vide.

Pour pouvoir appliquer les résultats qui vont suivre, il faut compactifier l'espace E . Soit $E' = E \cup \{\delta\}$ le compactifié d'ALEXANDROFF de E . On munit E' d'une métrique d' compatible avec sa topologie. E' est alors complet.

Soient P_0 une probabilité sur E' qui ne charge que E ($P_0\{\delta\} = 0$) et \mathfrak{M}' l'ensemble des probabilités sur E' . On définit comme en iii)

$v_1(A) = (1 - \epsilon) P_0(A) + \epsilon$	si $A \neq \emptyset$	$v_1(\emptyset) = 0$
$v_2(A) = \inf (P_0(A) + \epsilon, 1)$	si $A \neq \emptyset$	$v_2(\emptyset) = 0$
$v_3(A) = \inf (P_0(A^\delta) + \epsilon, 1)$	si $A \neq \emptyset$	$v_3(\emptyset) = 0$

$$A^\delta = \{x \in E' \mid d'(x, A) \leq \delta\}$$

$$\varphi_{v_1} = \{P = (1 - \epsilon) P_0 + \epsilon H' \mid H' \in \mathcal{H}'\}$$

$$\varphi_{v_2} = \{P \in \mathcal{H}' \mid \|P_0 - P\| \leq \epsilon\}$$

$$\varphi_{v_3} = \{P \in \mathcal{H}' \mid P(A) \leq P_0(A^\delta) + \epsilon\}$$

Les fonctions v_i sont bien des capacités mais les ensembles φ_{v_i} sont maintenant des voisinages de P_0 dans \mathcal{H}' et contiennent des mesures qui chargent le point à l'infini et sont sous-stochastiques sur E .

5 - LE LEMME DE NEYMAN-PEARSON POUR DES CAPACITÉS.

Soient v_0, v_1 deux capacités bi-alternatives, u_0, u_1 leurs conjuguées. On posera $\vartheta_{v_0} = \vartheta_0$ et $\vartheta_{v_1} = \vartheta_1$.

Soit un test de ϑ_0 contre ϑ_1 qui rejette ϑ_0 si x est dans A ; les probabilités maximales d'erreurs sont respectivement :

$$v_0(A) = \sup_{P \in \vartheta_0} P(A) \quad \text{si on rejette à tort } \vartheta_0$$

$$v_1(A^c) = 1 - u_1(A) = 1 - \inf_{P \in \vartheta_1} P(A) \quad \text{si on rejette } \vartheta_1 \text{ à tort.}$$

a) Le problème Bayésien :

On cherche des régions de rejet optimales pour le cas Bayésien, c'est-à-dire une famille d'ensemble A_t qui minimisent le risque Bayésien maximal:

$$A_t \text{ minimise } \frac{t}{1+t} v_0(A) + \frac{1}{1+t} (1 - u_1(A)) \quad 0 \leq t \leq \infty$$

A_t s'obtient en minimisant la fonction

$$w_t(A) = t v_0(A) - u_1(A) \quad 0 \leq t \leq \infty \quad w_\infty(A) = v_0(A)$$

Remarque : w_t n'est pas une capacité puisque elle n'est pas monotone. Mais elle est bi-alternative et possède la propriété de semi-continuité inférieure suivante :

Si (B_n) est une suite monotone d'ensembles qui converge vers B , on a d'après les propriétés (c) de v_0 et (c') de u_1 :

$$(f) \quad w_t(B) \leq \lim_n w_t(B_n)$$

PROPOSITION (5.1) :

Pour tout t dans $[0, +\infty]$, on peut trouver $A_t \in \mathcal{A}$ tel que :

$$w_t(A_t) = c_t = \inf_{A \in \mathcal{A}} w_t(A) .$$

Preuve :

Soient t fixé et (A_n) une suite d'ensembles tels que $w_t(A_n) \leq c_t + \epsilon_n$ avec $\epsilon_n > 0 \quad \forall n$ et $\sum_{n=1}^{\infty} \epsilon_n < +\infty$. On montre alors (voir détail des calculs dans [51]) que $\overline{\lim} A_n$ et $\underline{\lim} A_n$ minimisent w .

LEMME (5.2) :

Les ensembles A_t peuvent être choisis de façon à former une famille décroissante continue à droite c'est-à-dire

$$A_t = \bigcup_{s > t} A_s \quad \text{si } t \in [0, +\infty[\quad A_\infty = \bigcap_{t < \infty} A_t$$

Preuve :

Pour $0 \leq t < s < +\infty$, $w_t - w_s = (t - s) v_0$ est une fonction décroissante sur \mathcal{A} .

On a donc

$$w_t(A \cup B) - w_s(A \cup B) \leq w_t(A) - w_s(A)$$

$$w_s(A \cup B) + w_s(A \cap B) \leq w_s(A) + w_s(B)$$

soit
$$w_t(A \cup B) + w_s(A \cap B) \leq w_t(A) + w_s(B)$$

Soit $(A'_t)_{t \in \mathbb{R}^+}$ une famille telle que A'_t minimise w_t . On aura donc

$$c_s + c_t \leq w_t(A'_t \cup A'_s) + w_s(A'_t \cap A'_s) \leq w_t(A'_t) + w_s(A'_s) = c_s + c_t$$

qui prouve que $A'_t \cup A'_s$ minimise w_t .

Soit alors (t_n) une suite dense dans $[0, +\infty[$, si $N \geq n$, $\bigcup_{\substack{t_m \geq t_n \\ m \leq N}} A'_{t_m}$ minimise

aussi w_{t_n} et par semi-continuité de w $A''_{t_n} = \bigcup_{t_m \geq t_n} A'_{t_m}$ minimise aussi w_{t_n} .

Pour $t \in [0, +\infty[$, on construit une famille décroissante (A_t) en posant

$A_t = \bigcup_{t_n > t} A'_{t_n}$. On a $A_t = \bigcup_{s > t} A_s$. De plus si $(t_{n_k})_{k \in \mathbb{N}}$ est une sous-suite

décroissant vers t , $t_{n_k} \neq t$, $\forall k$ et $A''_{t_{n_k}}$ tend vers A_t .

$$c_t \leq w_t(A_t) \leq \lim_{t_{n_k} \rightarrow t} w_t(A''_{t_{n_k}}) \leq \lim_{t_{n_k} \rightarrow t} [w_{t_{n_k}}(A''_{t_{n_k}}) + (t_{n_k} - t)] = \lim_{t_{n_k} \rightarrow t} c_{t_{n_k}}$$

car $w_t \leq w_{t_{n_k}} + (t_{n_k} - t) v_0$.

Comme $|w_{t_{n_k}} - w_t| \leq t_{n_k} - t$, $|c_{t_{n_k}} - c_t| \leq t_{n_k} - t$ et c_t est une fonction

continue de t , ce qui entraîne $w_t(A_t) = c_t$. On pose alors $A_\infty = \bigcap_{t \in \mathbb{R}^+} A_t$. Comme $w_t(A_t) \leq w_t(\phi) = 0$, on a $v_0(A_t) \leq \frac{1}{t}$ soit $v_0(A_\infty) = 0$ ce qui prouve que A_∞ minimise $w_\infty = v$.

b) Dérivée de RADON-NIKODYM de v_1 par rapport à v_0 :

On a obtenu une famille (A_t) de régions de rejet optimales. Pour que celles-ci correspondent à des tests de rapport de vraisemblance, il faut que la fonction de vraisemblance γ vérifie $A_t = \{\gamma > t\}$. Ceci est réalisé si on choisit

$$\gamma(x) = \inf \{t \mid x \notin A_t\}$$

qui sera dite dérivée de RADON-NIKODYM de v_1 par rapport à v_0 . En effet si $t < s$ et $A_t \supset B \supset A_s$, on a :

$$w_t(B) \geq w_t(A_t) \Rightarrow t [v_0(B) - v_0(A_t)] \geq u_1(B) - u_1(A_t)$$

$$w_s(B) \geq w_s(A_s) \Rightarrow s [v_0(B) - v_0(A_s)] \geq u_1(B) - u_1(A_s)$$

Si les quotients ont un sens, on obtient les relations

$$(2) \frac{u_1(B) - u_1(A_s)}{v_0(B) - v_0(A_s)} \leq s ; \quad (3) \frac{u_1(A_t) - u_1(B)}{v_0(A_t) - v_0(B)} \geq t ; \quad (4) \quad t \leq \frac{u_1(A_t) - u_1(A_s)}{v_0(A_t) - v_0(A_s)} \leq s$$

Ces inégalités correspondent aux propriétés de la dérivée de RADON-NIKODYM :

Supposons que v_0 et v_1 soient des probabilités, alors $u_1 = v_1$, $v_0(A_\infty) = 0$ et $u_1(A_0) = 1$ donc

$$v_0 + v_1 \text{ p.s. } \frac{d v_1}{d v_0} = \begin{cases} \infty & \text{sur } A_\infty \\ 0 & \text{sur } A_0^c \end{cases} \quad \text{comme } \gamma$$

De plus les relations (2), (3), (4) impliquent que v_1 est absolument continue par rapport à v_0 sur $A_t - A_s$ avec $t \leq \frac{d v_1}{d v_0} \leq s$ v_0 p.s. et sur A_t $\frac{d v_1}{d v_0} \geq t$ v_0 p.s.

S'il existe un nombre t tel que l'on ait $\frac{d v_1}{d v_0}(x) < t < \gamma(x)$, x est dans A_t ce qui est contradictoire avec l'inégalité précédente. On a donc $\frac{d v_1}{d v_0} \geq \gamma$ v_0 p.s.

Sur A_t^c , $\frac{d v_1}{d v_0}$ est inférieur à t v_0 p.s. et on montre de même que $\frac{d v_1}{d v_0} \leq \gamma$ v_0 p.s.

On a donc $\gamma = \frac{d v_1}{d v_0}$ v_0 p.s. ce qui justifie sa dénomination. Le résultat était prévisible puisque (A_t) forme une famille de régions de rejet optimales pour les tests de v_0 contre v_1 et doivent donc être de la forme $\{\frac{d v_1}{d v_0} > t\}$ d'après le lemme de NEYMAN-PEARSON. Ce qui précède en constitue donc une nouvelle démonstration.

Remarque : En vue d'utiliser le théorème (1.3) sur les tests séquentiels, on peut chercher des conditions suffisantes pour que $\text{Log } \gamma$ soit borné :

- si $v_0(A) \geq \epsilon_0 > 0$ pour $A \neq \emptyset$, on a $v_0(A_t) \leq \frac{1}{t}$ et donc A_t est vide dès que t est supérieure à $\frac{1}{\epsilon_0}$. Dans ce cas γ est majoré par $\frac{1}{\epsilon_0}$.

- si $v_1(A) > \epsilon_1 > 0$ pour $A \neq \emptyset$, on a aussi $u_1(A) \leq 1 - \epsilon_1$ si $A \neq E$ et $w_t(E) = t - 1$.

Donc $w_t(A_t) = t v_0(A_t) - u_1(A_t) \leq t - 1$, soit $u_1(A_t) \geq 1 - t$.

Si t est inférieur à ϵ_1 , $A_t = E$ et γ est minoré par ϵ_1 .

Si donc ces deux conditions sont remplies, comme par exemple dans (4-iii), $\text{Log } \gamma$ sera borné.

c) Un couple minimal :

Puisque on souhaite que les régions de rejet $\{\gamma > t\}$ correspondent à celles de tests de rapport de vraisemblance, on va chercher un couple (Q_0, Q_1) qui vérifie :

$$\frac{d Q_1}{d Q_0} = \gamma (Q_0 + Q_1) \text{ p.s.} \quad Q_0 \leq v_0 \quad Q_1 \leq v_1$$

Pour cela on va construire Q_0 de façon que l'on ait $\frac{1}{\gamma} u_1 \leq Q_0 \leq v_0$ en définissant $\frac{1}{\gamma} u_1$ par des méthodes semblables à celles des intégrales de STIELTJES où les intervalles $[t, s[$ seraient remplacés par $A_t - A_s$.

THÉORÈME (5.3) :

Soient (E, \mathcal{A}) un espace séparable, métrique, complet, muni de la tribu de ses boréliens, v_0 et v_1 deux capacités bi-alternatives sur E , γ la dérivée de RADON-NIKODYM de v_1 par rapport à v_0 . Il existe deux probabilités $Q_0 \leq v_0$ et $Q_1 \leq v_1$, telles que pour tout t on ait :

$$Q_0 \{\gamma > t\} = v_0 \{\gamma > t\} \quad Q_1 \{\gamma > t\} = u_1 \{\gamma > t\}$$

$$\gamma = \frac{d Q_1}{d Q_0} (Q_0 + Q_1) \text{ p.s.}$$

Les tests de Q_0 contre Q_1 constituent donc des tests minimax de ϑ_{v_0} contre ϑ_{v_1} .

Preuve : Pour $t < s$ et B dans \mathcal{A} on pose :

$$G(t, s; B) = v_0((B \cap A_t) \cup A_s) - v_0(A_s)$$

Pour C compact :

$$(5) \quad F(t, s; C) = u_1((C \cap A_t) \cup A_s) - u_1(A_s)$$

$$\text{et} \quad F(t, s; B) = \sup_C F(t, s; C) \quad C \text{ compact inclus dans } B.$$

G est monotone et bi-alternative, F monotone et bi-continue, F vérifie (d') parce qu'un compact inclus dans une réunion croissante d'ouverts est inclus dans l'un d'eux.

D'après [18] u_1 vérifie la propriété :

$$K_i \subset A_i \quad 1 \leq i \leq n \quad \Rightarrow \quad u(\cap A_i) - u(\cap K_i) \leq \sum_i [u(A_i) - u(K_i)]$$

Ceci permet de voir que F possède également la propriété (c'). De plus d'après (2) on a :

$$(6) \quad F(t, s; B) \leq s G(t, s; B)$$

En appliquant (e) aux ensembles $(B \cap A_t)$ et A_s on montre :

$$(7) \quad G(t, s; B) \leq v_o(B \cap A_t) - v_o(B \cap A_s) \leq v_o(B)$$

De plus pour $t_1 < t_2 < t_3$ on a d'après (e) et (e')

$$(8) \quad G(t_1, t_2; B) + G(t_2, t_3; B) \leq G(t_1, t_3; B)$$

$$(9) \quad F(t_1, t_2; B) + F(t_2, t_3; B) \geq F(t_1, t_3; B)$$

- On montre que si $B = A_r$ ou $B = C \cap A_r$ où C est compact, l'égalité (5) demeure vraie.

Si $B = C \cap A_r \quad t \leq r \leq s$ et K est un compact inclus dans A_r

$$\begin{aligned} F(t, s; B) &\geq \sup_K u_1((C \cap K) \cup A_s) - u_1(A_s) = \sup_K u_1((C \cup A_s) \cap (K \cup A_s)) - u_1(A_s) \\ &\geq u_1((C \cup A_s) \cap A_r) - u_1(A_s) = u_1((C \cap A_r) \cup A_s) - u_1(A_s) \end{aligned}$$

puisque les boréliens sont u_1 -capacitables.

Comme par définition $F(t, s; B) \leq u_1((C \cap A_r) \cup A_s) - u_1(A_s)$, l'égalité (5) est vérifiée. Les autres cas sont analogues.

- Si $J = \{t_0, t_1, \dots, t_n\}$ est une famille finie de réels positifs, on pose

$$u_J(B) = \sum_{i=1}^n \frac{1}{t_i} F(t_{i-1}, t_i; B)$$

Ceci va permettre de définir une fonction u^* sur \mathcal{A} , telle que si u_1 est une probabilité, on ait $\frac{du^*}{du_1} = \frac{1}{\gamma}$.

D'après (6), (8) et (7) on a

$$u_J(B) \leq \sum_{i=1}^n G(t_{i-1}, t_i; B) \leq G(t_0, t_n; B) \leq v_0(B)$$

et si J' est une subdivision plus fine que J ($J' \supset J$) on a

$$u_{J'}(B) \geq u_J(B) \quad \text{par (9)}$$

Ceci permet de définir $u(B) = \sup_J u_J(B) \leq v_0(B)$, u , comme F , est positive, monotone et bi-monotone. Calculons $u(A_t)$:

$$\begin{aligned} G(t_1, t_2; A_t) &= v_0(A_{t_1}) - v_0(A_{t_2}) && \text{si } t \leq t_1 \\ &= v_0(A_t) - v_0(A_{t_2}) && \text{si } t_1 < t < t_2 \\ &= 0 && \text{si } t \geq t_2 \end{aligned}$$

F vérifie des formules analogues avec u_1 à la place de v_0 ce qui implique d'après (4) :

$$(10) \quad F(t_1, t_2; A_t) \geq t_1 G(t_1, t_2; A_t)$$

Choisissons alors une partition J assez fine pour que $\frac{t_{i-1}}{t_i} \geq \alpha$ pour tout i et $t = t_{i_0}$ pour un certain i_0 ; on obtient [51]

$$u_J(A_t) \geq \alpha [v_0(A_t) - v_0(A_{t_n})] \quad \text{avec} \quad v_0(A_{t_n}) \leq \frac{1}{t_n}$$

Si on fait tendre α vers 1 et t_n vers l'infini en augmentant le nombre des points de subdivision, on conclut que $u(A_t) \geq v_0(A_t)$ d'où l'égalité

$$u(A_t) = v_0(A_t) \quad \text{pour} \quad 0 < t < +\infty$$

Par monotonie :

$$u(A_0) \geq \lim_{t \rightarrow 0} u(A_t) = v_0(A_0)$$

et

$$0 \leq u(A_\infty) \leq u(A_n) = v_0(A_n) \leq \frac{1}{n}$$

ce qui montre que l'égalité demeure vraie pour $t = 0$ ou $+\infty$.

Comme $F(t, s; E) = F(t, s; A_0)$ on a $u(E) = u(A_0) = v_0(A_0)$.

Si $v_0(A_0)$ est plus petit que 1, on normalise u en le remplaçant par $u^* = u + Q_{|A_0^c}$ où Q est une probabilité majorée par v_0 telle que $Q(A_0) = v_0(A_0)$ (cf. Lemme (3.5)). On a alors $u^*(E) = 1$.

$$\begin{aligned} u^*(B) &= u(B \cap A_0) + Q(B \cap A_0^c) = u(B \cap A_0) + Q(B \cup A_0) - Q(A_0) \\ &\leq v_0(B \cap A_0) + v_0(B \cup A_0) - v_0(A_0) \leq v_0(B) \end{aligned}$$

u^* est majorée par v_0 comme u .

- Construisons Q_0 :

On considère les ensembles disjoints

$$U = \{g \in \mathcal{C} \mid \tilde{v}_0(|g|) < 1\} \quad \text{ouvert}$$

$$V = \{f \in \mathcal{C}^+ \mid \tilde{u}^*(f) \geq 1\} \quad \text{fermé}$$

Ce sont deux ensembles convexes car \tilde{v}_0 est sous-additive et \tilde{u}^* sur-additive.

On applique aux convexes U et V le théorème de HAHN-BANACH et la démonstration de la proposition (3.4). On en déduit l'existence d'une probabilité Q_0 vérifiant

$$\tilde{u}^*(f) \leq \tilde{Q}_0(f) \leq \tilde{v}_0(f) \quad \text{pour } f \text{ dans } \mathcal{C}^+.$$

On a aussi $\tilde{Q}_0 \geq \tilde{u}_J$ sur \mathcal{C}^+ . Puisque u_J possède comme F les propriétés (d) et de capacabilité, on peut lui appliquer le lemme (2.5), qui implique que $Q_0 \geq u_J$.

On a donc :

$$Q_0(A_t) \geq u_J(A_t) \geq \alpha [v_0(A_t) - v_0(A_{t_n})]$$

Par passage à la limite, il vient $Q_0(A_t) \geq v_0(A_t)$ et donc $Q_0(A_t) = v_0(A_t)$.
En particulier $Q_0(A_\infty) = 0$.

- On définit alors Q_1 comme suit :

Soit Q dans \mathcal{P}_1 telle que $Q(A_\infty) = u_1(A_\infty)$, on pose $Q_1|_{A_\infty} = Q|_{A_\infty}$ de sorte que sur A_∞ on ait $\frac{dQ_1}{dQ_0} = +\infty = \gamma(Q_1 + Q_0)$ p.s.

Sur A_∞^c on choisit Q_1 de manière à avoir $dQ_1 = \gamma dQ_0$. Il reste à vérifier que Q_1 est une probabilité minorée par u_1 , telle que $Q_1(A_t) = u_1(A_t)$ pour tout t .

Soient C un compact, $\alpha < 1$, $J = \{t_0, \dots, t_n\}$ avec $\frac{t_{k-1} - t_0}{t_k} \geq \alpha$ si $k > 1$.
On a alors ([51])

$$\begin{aligned} Q_1(C \cap A_0) - Q_1(C \cap A_\infty) &\geq \sum_{1 \leq i < k \leq n} \frac{t_i - t_{i-1}}{t_k} F(t_{k-1}, t_k; C \cap A_k) \\ &= \sum_{1 \leq i < k \leq n} \frac{t_i - t_{i-1}}{t_k} F(t_{k-1}, t_k; C) \end{aligned}$$

puisque l'égalité (5) est vérifiée.

$$\geq \alpha [u_1(C \cap A_{t_1}) - u_1(C \cap A_{t_n})]$$

En utilisant (e') et (2) on a aussi

$$\begin{aligned} u_1(C \cap A_0) - u_1(C \cap A_t) &\leq u_1((C \cap A_0) \cup A_t) - u_1(A_t) \leq t [v_0((C \cap A_0) \cup A_t) \\ &\quad - v_0(A_t)] \leq t \end{aligned}$$

qui implique la continuité de $u_1(C \cap A_t)$ lorsque t tend vers 0. Alors, si on fait tendre α vers 1, t_0 et t_1 vers 0 et t_n vers l'infini, on obtient à la limite

$$Q_1(C \cap A_0) - Q_1(C \cap A_\infty) \geq u_1(C \cap A_0) - u_1(C \cap A_\infty)$$

Comme $1 \geq u_1(C \cup A_0) \geq u_1(A_0) = 1$ on peut écrire

$$u_1(C \cap A_0) + u_1(C \cup A_0) \geq u_1(C) + u_1(A_0) = 1 + u_1(C)$$

ce qui implique $u_1(C \cap A_0) = u_1(C)$

Ceci permet d'écrire :

$$Q_1(C) \geq Q_1(C \cap A_0) \geq u_1(C) + Q_1(C \cap A_\infty) - u_1(C \cap A_\infty) \geq u_1(C)$$

puisque sur A_∞ on a pris $Q_1 = Q \geq u_1$.

Par capacabilité, cette inégalité est vraie pour tous les ensembles B de \mathcal{A} .

Si t_1 est inférieur à t_2 , on a

$$\begin{aligned} Q_1(A_{t_1}) - Q_1(A_{t_2}) &\leq t_2 [Q_0(A_{t_1}) - Q_0(A_{t_2})] = t_2 [v_0(A_{t_1}) - v_0(A_{t_2})] \\ &\leq t_2/t_1 [u_1(A_{t_1}) - u_1(A_{t_2})] \quad \text{d'après (3).} \end{aligned}$$

En choisissant une famille $\{t_i\}$ telle que l'on ait :

$$0 < t = t_0 < t_1 < \dots < t_n \quad \text{avec} \quad \frac{t_i}{t_i - 1} \quad \text{et} \quad \alpha > 1,$$

on peut écrire

$$Q_1(A_t) - Q_1(A_{t_n}) \leq \alpha [u_1(A_t) - u_1(A_{t_n})]$$

Par passage à la limite si α tend vers 1 et t_n vers l'infini

$$Q_1(A_t) - Q_1(A_\infty) \leq u_1(A_t) - u_1(A_\infty)$$

Comme $Q_1(A_\infty) = u_1(A_\infty)$, il en résulte que $Q_1(A_t) = u_1(A_t)$, pour t non nul.

Et comme $u_1(A_t) \geq 1 - t$ tend vers $u_1(A_0)$ si t tend vers 0, le résultat demeure vrai si t est nul par continuité.

$$\text{On a donc } Q_1(A_0) = 1 \quad \text{et} \quad Q_1(A_0^c) = \int 1_{A_0^c} \gamma(x) dQ_0 = 0$$

puisque $\gamma = 0$ sur A_0^c .

Ceci montre que Q_1 est bien une probabilité.

C.Q.F.D.

d) Cas particulier d'un compactifié d'ALEXANDROFF :

On se place dans le cadre des exemples du (4-iv). L'espace $E' = E \cup \{\delta\}$ considéré est le compactifié d'ALEXANDROFF d'un espace E et les capacités v_0 et v_1 vérifient les propriétés suivantes :

$$(*) \quad v_i(\{\delta\}) = \epsilon_i \geq 0 \quad v_i(A \cup \{\delta\}) = v_i(A) \quad \text{si } A \neq \emptyset \quad i = 0, 1.$$

Ces propriétés sont celles des capacités v_1, v_2, v_3 considérées en (4-iv), c'est-à-dire celles des enveloppes supérieures de voisinages d'une probabilité P qui ne charge pas le point à l'infini. Dans ce cas on a le résultat :

COROLLAIRE (5.4) :

Sous les hypothèses (*) précédentes, le couple (Q_0, Q_1) minimal du théorème (5.3) est tel que ni Q_0 , ni Q_1 ne chargent δ ; leur support est inclus dans E .

Remarque :

Ce corollaire permet d'appliquer les résultats généraux aux espaces \mathbb{R}^n : même si l'on considère des voisinages dans le compactifié, si leurs centres sont des probabilités sur \mathbb{R}^n , le couple minimal est un couple de probabilités sur \mathbb{R}^n .

Preuve :

On a $u_1(E) = 1 - \epsilon_1$ et $u_1(A \cup \{\delta\}) = u_1(A)$ si $A \neq E$
en particulier $u_1(\{\delta\}) = u_1(\emptyset) = 0$.

- Comme $v_0(A_\infty)$ est nul, si δ est dans A_∞ , c'est que ϵ_0 est nul. v_0 ne charge pas δ , Q_0 non plus et on a

$$Q_1(A_\infty - \delta) \geq u_1(A_\infty - \delta) = u_1(A_\infty) = Q_1(A_\infty) \quad \text{donc} \quad Q_1(\{\delta\}) = 0.$$

- Si $\gamma(\delta)$ est nul, δ appartient à A_0^c . Comme $v_1(A_0^c) = 0$, ϵ_1 est nul et Q_1 ne charge pas δ .

$$Q_0(A_0 \cup \{\delta\}) \leq v_0(A_0 \cup \{\delta\}) = v_0(A_0) = Q_0(A_0) \quad \text{donc } Q_0(\{\delta\}) = 0$$

- On étudie le cas général $0 < \gamma(\delta) < +\infty$.

S'il existe un t tel que $\delta \in A_t$ et $A_t \neq E'$ on a

$$Q_1(A_t - \delta) \geq u_1(A_t - \delta) = u_1(A_t) = Q_1(A_t)$$

$$\text{et donc} \quad 0 = Q_1(\delta) = Q_0(\delta) \gamma(\delta) \quad \text{d'où} \quad Q_1(\delta) = Q_0(\delta) = 0.$$

Dans le cas contraire, tout A_t contenant δ est égal à E' et donc pour $t < \gamma(\delta)$ $A_t = E'$.

$A_\gamma(\delta)$ ne contient pas δ . Supposons qu'il soit vide, on aura $c_t = t - 1$ pour $t < \gamma(\delta)$, $c_t = 0$ pour $t \geq \gamma(\delta)$. Comme c_t est continu ceci implique que $\gamma(\delta) = 1$ et donc $A_t = E'$ si $t < 1$, $A_t = \emptyset$ si $t \geq 1$ soit $\gamma(x) = 1$ pour tout x dans E' ; comme A_∞ est vide, on peut dans la démonstration du théorème (5.3) prendre Q_1 égale à Q_0 . Mais alors les ensembles \mathcal{C}_{v_0} et \mathcal{C}_{v_1} ne sont plus disjoints ce qui est contraire aux hypothèses.

$A_\gamma(\delta)$ n'étant pas vide, on peut écrire :

$$Q_0(A_\gamma(\delta) \cup \{\delta\}) \leq v_0(A_\gamma(\delta) \cup \{\delta\}) = v_0(A_\gamma(\delta)) = Q_0(A_\gamma(\delta))$$

et $Q_1(\delta) = \gamma(\delta) Q_0(\delta)$. Ni Q_0 , ni Q_1 ne chargent δ .

C.Q.F.D.

e) Unicité de γ :

On reprend les hypothèses du théorème (5.3).

PROPOSITION (5.5) :

Soient deux familles d'ensembles $A_t = \{\gamma > t\}$ et $A'_t = \{\gamma' > t\}$ qui minimisent w_t , γ et γ' étant deux versions de $\frac{dv_1}{dv_0}$; on a :

$$v_0\{\gamma > t\} = v_0\{\gamma' > t\} \quad \text{et} \quad u_1\{\gamma > t\} = u_1\{\gamma' > t\} \quad \forall t.$$

De plus si Q_o (respectivement Q'_o) est la probabilité construite dans le théorème (5.3) à partir de γ (respectivement γ'), on a :

$$\gamma = \gamma' (Q_o + Q'_o) \quad \text{p.s.}$$

Preuve :

D'après l'inégalité

$$w_t (A_t \cup A'_t) + w_t (A_t \cap A'_t) \leq w_t (A_t) + w_t (A'_t)$$

les familles $A_t \cup A'_t$ et $A_t \cap A'_t$ minimisent également w_t . Elles sont associées à $\sup (\gamma, \gamma')$ et $\inf (\gamma, \gamma')$ respectivement puisque l'on a :

$$A_t \cup A'_t = \{\sup (\gamma, \gamma') > t\} \quad A_t \cap A'_t = \{\inf (\gamma, \gamma') > t\}$$

Donc $\sup (\gamma, \gamma')$ et $\inf (\gamma, \gamma')$ sont deux autres versions de $\frac{dv_1}{dv_o}$ ce qui permet de supposer que l'on a $\gamma \leq \gamma'$.

Supposons qu'il existe t dans $]0, +\infty[$ tel que $v_o \{\gamma > t\} < v_o \{\gamma' > t\}$ et donc $u_1 \{\gamma > t\} < u_1 \{\gamma' > t\}$ puisque $t v_o (A_t) - u_1 (A_t) = t v_o (A'_t) - u_1 (A'_t)$. Si $D = A'_t - A_t$, on aura :

$$t = \frac{u_1 (A'_t) - u_1 (A_t)}{v_o (A'_t) - v_o (A_t)} \geq \frac{Q'_1 (A'_t) - Q'_1 (A_t)}{Q'_o (A'_t) - Q'_o (A_t)} \geq \frac{\int_D \gamma' dQ_o}{\int_D dQ'_o}$$

puisque on a $Q'_1 \geq u_1$, $Q'_o \leq v_o$, $Q'_1 (A'_t) = u_1 (A'_t)$, $Q'_o (A'_t) = v_o (A'_t)$

$$Q'_o (D) = Q'_o (A'_t) - Q'_o (A_t) \geq v_o (A'_t) - v_o (A_t) > 0.$$

De plus sur D , γ' est supérieur à t , donc le rapport de ces deux intégrales doit être plus grand que t , ce qui est impossible. Les égalités sont vraies pour $t = 0$ ou $t = +\infty$ par continuité.

Si l'on considère la construction du théorème (5.3) utilisant γ' et la famille A'_t , on a défini

$$F' (t, s ; B) = \sup_{K \subset B} u_1 ((K \cap A'_t) \cup A'_s) - u_1 (A'_s) \quad K \text{ compact.}$$

On va montrer que l'on a $F^1(t, s; A'_r) = F^1(t, s; A_r)$ pour tout r . A_r étant inclus dans A'_r , une inégalité est évidente. On considère 3 cas :

- $t < s \leq r$ on a alors

$$A_r \subset A'_r \subset A'_s \subset A'_t \text{ et l'égalité se réduit à}$$

$$F^1(t, s; A_r) = 0 = F^1(t, s; A'_r)$$

- $t < r < s$

$$u_1((A'_t \cap K) \cup A'_s) = u_1(K \cup A'_s) \text{ si } K \text{ est inclus dans } A_r \text{ ou } A'_r$$

et par u_1 -capacitabilité des boréliens il vient

$$\begin{aligned} u_1(A_r) &= \sup_{K \subset A_r} u_1(K) \leq \sup_{K \subset A_r} u_1(K \cup A'_s) \leq \sup_{K' \subset A'_r} u_1(K' \cup A'_s) \leq (A'_r \cup A'_s) \\ &= u_1(A'_r) \end{aligned}$$

L'égalité des termes extrêmes entraîne l'égalité cherchée.

- $r \leq t < s$

On a alors $A'_t \subset A'_r$, $A_t \subset A_r$ et donc

$$\sup_{K' \subset A'_r} u_1((A'_t \cap K') \cup A'_s) \leq u_1(A'_t)$$

$$\sup_{K \subset A_r} u_1((A'_t \cap K) \cup A'_s) \geq \sup_{K \subset A_t} u_1(A'_t \cap K) = u_1(A_t)$$

Comme $u_1(A'_t) = u_1(A_t)$, on a $F^1(t, s; A_r) = F^1(t, s; A'_r)$

Ceci permet de voir que l'on a également

$$u_1^*(A'_t) = u_1^*(A_t) \quad \forall t.$$

On peut alors vérifier que $\gamma = \gamma^1(Q_0 + Q'_0)$ p.s. puisque on a

$$v_o \{ \gamma > t \} = Q_o \{ \gamma > t \} \leq Q_o \{ \gamma' > t \} \leq v_o \{ \gamma' > t \}$$

$$u'^* \{ \gamma' > t \} = Q'_o \{ \gamma' > t \} \geq Q'_o \{ \gamma > t \} \geq u'^* \{ \gamma > t \}$$

6 - LIEN AVEC LA THÉORIE DES EXPÉRIENCES.

Les résultats précédents peuvent se traduire dans le langage de la théorie des expériences de BLACKWELL [11] et LE CAM [65], [67].

a) Rappels et définitions :

Une expérience \mathcal{E} est la donnée d'un espace mesuré (E, \mathcal{A}) , d'un espace de paramètres Θ , qu'on supposera fini égal à $\{0; 1; \dots; n\}$ et d'une famille $(P_\theta)_{\theta \in \Theta}$ de probabilités sur E . Si $\Theta = \{0; 1\}$ l'expérience est dite binaire.

L'espace D des décisions est un sous-ensemble borné de \mathbb{R}^n . Si $d = (d_0, \dots, d_n)$ est une décision, d_i représente la perte lorsque l'on prend la décision d et que la véritable distribution est P_i .

Une procédure de décision est une application f de E dans D :

$$f(x) = (f_0(x), \dots, f_n(x)) \in D$$

et on a un vecteur de perte

$$v(f) = \left(\int f_0(x) dP_0(x), \dots, \int f_n(x) dP_n(x) \right) \in \mathbb{R}^n$$

L'enveloppe convexe fermée de l'ensemble des $v(f)$ est appelée l'enveloppe des risque $R(\mathcal{E}, D)$. L'expérience \mathcal{E} sera dite meilleure que l'expérience \mathcal{F} ($\mathcal{E} > \mathcal{F}$), si

$$R(\mathcal{E}, D) \supset R(\mathcal{F}, D)$$

BLACKWELL a montré le résultat suivant [11].

THÉORÈME (6.1) :

Soient deux expériences $\mathcal{E} = (E, \mathcal{A}; P_0, \dots, P_n)$ et $\mathcal{F} = (F, \mathcal{B}; Q_0, \dots, Q_n)$.
Considérons les mesures $M = \sum_{i=0}^n P_i$ et $N = \sum_{i=0}^n Q_i$ et le simplexe S de $[0, 1]^{n+1}$:

$$S = \{(p_i) \mid p_0 + p_1 + \dots + p_n = 1\}$$

$\mathcal{S} > \mathcal{T}$ si et seulement si pour toute fonction g continue convexe sur S , on a :

$$\int g \left(\frac{dP_0}{dM}, \frac{dP_1}{dM}, \dots, \frac{dP_n}{dM} \right) dM \geq \int g \left(\frac{dQ_0}{dM}, \frac{dQ_1}{dM}, \dots, \frac{dQ_n}{dM} \right) dM$$

b) Expériences binaires minimales :

La relation $>$ définit sur l'ensemble des expériences un ordre partiel. Nous allons voir que si l'on considère l'ensemble des expériences binaires sur (E, \mathcal{A}) formées par des couples (P_0, P_1) , où $P_0 \in \mathcal{P}_{v_0}$, $P_1 \in \mathcal{P}_{v_1}$, v_0 et v_1 étant des capacités bi-alternatives, l'expérience $\mathcal{S} = (E, \mathcal{A}; Q_0, Q_1)$, où (Q_0, Q_1) est le couple minimal du théorème (5.3) est minimale pour la relation d'ordre $>$, ce qui s'écrit :

$$\forall \mathcal{T} = (E, \mathcal{A}; P_0, P_1) \quad , \quad P_0 \in \mathcal{P}_{v_0}, P_1 \in \mathcal{P}_{v_1} \quad \mathcal{S} < \mathcal{T}$$

Pour obtenir ce résultat, on va démontrer la condition suffisante du théorème (6.1). Dans le cas d'expériences binaires, on posera $g(x, 1-x) = \Phi(x)$. Φ est une fonction continue, convexe sur $[0, 1]$. On pourra supposer que Φ est de classe \mathcal{C}^2 puisque \mathcal{C}^2 est dense dans \mathcal{C}^0 pour la topologie de la convergence uniforme. De plus, quitte à lui ajouter une fonction linéaire $f(x) = ax + b$, ce qui revient à ajouter une même constante $\int [a \frac{dP_0}{dM} + b] dM = a + 2b$ à toutes les intégrales, on peut supposer que l'on a $\Phi(0) = \Phi'(0) = 0$, ce qui implique que Φ et Φ' sont positifs sur $[0, 1]$, Φ étant convexe.

THÉORÈME (6.2) :

Soit Φ une fonction convexe, positive, croissante, de classe \mathcal{C}^2 sur $[0, 1]$, telle que $\Phi(0) = \Phi'(0) = 0$. Sous les hypothèses du théorème (5.3), le couple (Q_0, Q_1) construit dans ce théorème minimise la fonction

$$H(P_0, P_1) = \int \Phi \left(\frac{dP_0}{d(P_0 + P_1)} \right) d(P_0 + P_1)$$

sur l'ensemble $\mathcal{P}_{v_0} \times \mathcal{P}_{v_1}$.

Remarque :

L'idée de la démonstration est celle de HUBER et STRASSEN dans [51], mais leur preuve est incomplète parce qu'ils ne considèrent que le cas où $Q_0 + Q_1$ domine $P_0 + P_1$. La réciproque qu'ils obtiennent ainsi doit être remplacée par le théorème (6.3).

Preuve :

Soit (P_0, P_1) un couple quelconque dans $\mathcal{P}_{V_0} \times \mathcal{P}_{V_1}$. On va d'abord construire un couple (Q'_0, Q'_1) tel que $H(Q_0, Q_1)$ et $H(Q'_0, Q'_1)$ soient très proches et que $(Q'_0 + Q'_1)$ domine $(P_0 + P_1)$.

- Soit γ la version de $\frac{dQ_1}{dQ_0}$ considérée dans le théorème (5.3). On a $A_\infty = \{\gamma = \infty\}$ et $Q_0(A_\infty) = 0$.

Posons $\mu = P_0 + P_1 + Q_0 + Q_1$ et soit θ une mesure telle que $d\theta = 1_{A_\infty^c} d\mu$.

Sur A_∞^c , γ est fini et il existe une fonction f sur E , avec

$$1 \geq f(x) > 0 \quad \forall x \in E \quad \text{et} \quad \int \gamma(x) f(x) d\theta < +\infty.$$

On note $(1 - \gamma)^+ = \sup \{(1 - \gamma), 0\}$ et $(1 - \gamma)^- = -\inf \{(1 - \gamma), 0\}$,

$$\alpha = \int (1 - \gamma)^+ f d\theta \quad \beta = \int (1 - \gamma)^- f d\theta$$

$$x = \int_{\{\gamma < 1\}} f d\theta \quad y = \int_{\{\gamma \geq 1\}} f d\theta$$

α et β sont strictement positifs, ainsi que x et y .

On considère alors une mesure T_0 ayant pour densité par rapport à θ la fonction $[a 1_{\{\gamma < 1\}} + b 1_{\{\gamma \geq 1\}}] f$, de sorte que l'on a

$$T_0(E) = ax + by \quad T_0(A_\infty) = 0$$

Soit ϵ_1 positif, on construit T_1 comme suit :

T_1 est une mesure égale à γT_0 sur A_∞^c et équivalente à μ sur A_∞ avec $T_1(A_\infty) = \epsilon_1$. On a alors :

$$\begin{aligned} T_1(A_\infty^c) &= a \int_{\{\gamma < 1\}} \gamma f d\theta + b \int_{\{\gamma \geq 1\}} \gamma f d\theta = ax + by - a \int_{\{\gamma < 1\}} (1-\gamma) f d\theta \\ &\quad - b \int_{\{\gamma \geq 1\}} (1-\gamma) f d\theta \\ &= a(x - \alpha) + b(y + \beta) \end{aligned}$$

On va choisir ϵ_1 , a et b positifs de façon que T_0 et T_1 soient deux probabilités.

Il suffit de prendre

$$a = \frac{\beta + \epsilon_1 y}{\alpha y + x \beta} \quad b = \frac{\alpha - x \epsilon_1}{\alpha y + x \beta}$$

qui seront positifs pourvu que ϵ_1 soit assez petit.

On a construit ainsi deux probabilités T_0 et T_1 vérifiant $\frac{dT_1}{dT_0} = \gamma \mu$ p.s. , $(T_0 + T_1)$ étant une mesure équivalente à μ .

Soit alors ϵ positif arbitrairement petit, on posera

$$Q'_0 = (1 - \epsilon) Q_0 + \epsilon T_0 \quad Q'_1 = (1 - \epsilon) Q_1 + \epsilon T_1$$

$(Q'_0 + Q'_1)$ est une mesure équivalente à μ et domine donc $(P_0 + P_1)$.

On a $\frac{dQ'_1}{dQ'_0} = \gamma \mu$ p.s. et

$$|H(Q_0, Q_1) - H(Q'_0, Q'_1)| = \left| \int \Phi\left(\frac{1}{1+\gamma}\right) d(Q_0 + Q_1 - Q'_0 - Q'_1) \right| \leq 4 \epsilon \|\Phi\|$$

$$\text{On pose} \quad p_0 = \frac{dP_0}{d\mu} \quad p_1 = \frac{dP_1}{d\mu} \quad q'_0 = \frac{dQ'_0}{d\mu} \quad q'_1 = \frac{dQ'_1}{d\mu}$$

$$\begin{aligned} R_{0t} &= (1-t) Q'_0 + t P_0 & r_{0t} &= \frac{dR_{0t}}{d\mu} \\ R_{1t} &= (1-t) Q'_1 + t P_1 & r_{1t} &= \frac{dR_{1t}}{d\mu} \end{aligned} \quad \left. \vphantom{\begin{aligned} R_{0t} \\ R_{1t} \end{aligned}} \right\} \forall t \in [0, 1]$$

$$I(t) = H(R_{ot}, R_{1t}) = \int \Phi\left(\frac{r_{ot}}{r_{ot} + r_{1t}}\right) (r_{ot} + r_{1t}) d\mu$$

si t est égal à 1, le dénominateur devient $p_o + p_1$ et peut être nul, on prend alors

l'expression $\Phi\left(\frac{r_{ot}}{r_{ot} + r_{1t}}\right) (r_{ot} + r_{1t})$ égale à 0.

Sur $[0, 1[$, $r_{ot} + r_{1t}$ est non nul μ p.s. puisque l'on a $(q'_o + q'_1) \neq 0$ μ p.s.

$I(t)$ est de classe \mathcal{C}^2 sur $[0, 1[$, continue en 1, et on obtient par dérivation sous le signe d'intégration les formules pour $t \in [0, 1[$.

$$I'(t) = \int \Phi\left(\frac{r_{ot}}{r_{ot} + r_{1t}}\right) (p_o + p_1 - q'_o - q'_1) d\mu + \int \Phi'\left(\frac{r_{ot}}{r_{ot} + r_{1t}}\right) \left(\frac{p_o q'_1 - q'_o p_1}{r_{ot} + r_{1t}}\right) d\mu$$

$$I''(t) = \int \Phi''\left(\frac{r_{ot}}{r_{ot} + r_{1t}}\right) \frac{(p_o q'_1 - p_1 q'_o)^2}{(r_{ot} + r_{1t})^3} d\mu \geq 0$$

La fonction I est donc convexe.

On considère les fonctions définies sur $[0, 1]$:

$$\varphi(z) = z \Phi'(z) - \Phi(z) \quad \psi(z) = (1 - z) \Phi'(z) + \Phi(z)$$

$$\varphi'(z) = z \Phi''(z) \quad \text{et} \quad \psi'(z) = (1 - z) \Phi''(z) \quad \text{sont positives.}$$

Comme $\varphi(0) = \psi(0) = 0$, φ et ψ sont deux fonctions positives, croissantes et bornées sur $[0, 1]$. On a les égalités

$$\begin{aligned} I'(0) &= \int \Phi\left(\frac{q'_o}{q'_o + q'_1}\right) (p_o + p_1 - q'_o - q'_1) d\mu + \int \Phi'\left(\frac{q'_o}{q'_o + q'_1}\right) \left(\frac{p_o q'_1 - p_1 q'_o}{q'_o + q'_1}\right) d\mu \\ &= \int \varphi\left(\frac{q'_o}{q'_o + q'_1}\right) (q'_1 - p_1) d\mu + \int \psi\left(\frac{q'_o}{q'_o + q'_1}\right) (p_o - q'_o) d\mu \\ &= \int \varphi\left(\frac{1}{1 + \gamma}\right) (q'_1 - p_1) d\mu + \int \psi\left(\frac{1}{1 + \gamma}\right) (p_o - q'_o) d\mu \end{aligned}$$

puisque $\frac{q'_1}{q'_o} = \gamma$ μ p.s.

En posant $\varphi\left(\frac{1}{1+z}\right) = \varphi_1(z)$ $\psi\left(\frac{1}{1+z}\right) = \psi_1(z)$ on obtient deux fonctions positives, décroissantes bornées sur $[0, +\infty]$ avec

$$I'(0) = \int \varphi_1(\gamma) (q'_1 - p_1) d\mu + \int \psi_1(\gamma) (p_0 - q'_0) d\mu$$

soit

$$I'(0) = \int Q'_1 \{\varphi_1 \circ \gamma \geq t\} dt - \int P_1 \{\varphi_1 \circ \gamma \geq t\} dt + \int P_0 \{\psi_1 \circ \gamma \geq t\} dt - \int Q'_0 \{\psi_1 \circ \gamma \geq t\} dt.$$

Si M est un majorant commun à φ_1 et ψ_1 , on peut écrire :

$$(11) \quad I'(0) \geq \int Q_1 \{\varphi_1 \circ \gamma \geq t\} dt - \int P_1 \{\varphi_1 \circ \gamma \geq t\} dt + \int P_0 \{\psi_1 \circ \gamma \geq t\} dt - \int Q_0 \{\psi_1 \circ \gamma \geq t\} dt - 4 \in M$$

Comme d'après le théorème (5.3) on a

$$Q_1 \{\gamma > t\} \leq P_1 \{\gamma > t\} \quad \text{et} \quad Q_0 \{\gamma > t\} \geq P_0 \{\gamma > t\}$$

et que φ_1 et ψ_1 sont monotones, on obtient les inégalités

$$(12) \quad Q_1 \{\varphi_1 \circ \gamma \geq t\} \geq P_1 \{\varphi_1 \circ \gamma \geq t\} \quad Q_0 \{\psi_1 \circ \gamma \geq t\} \leq P_0 \{\psi_1 \circ \gamma \geq t\}$$

Ceci montre que $I'(0)$ est supérieur à $-4 \in M$ et donc

$$I'(t) \geq -4 \in M \quad \forall t \in [0, 1[,$$

$$I(1) \geq I(0) - 4 \in M \quad \text{soit}$$

$$H(P_0, P_1) \geq H(Q'_0, Q'_1) - 4 \in M \geq H(Q_0, Q_1) - 4 \in M - 4 \in \|\phi\|$$

Comme ϵ est arbitraire, le résultat s'ensuit.

C.Q.F.D.

D'après le théorème de BLACKWELL, tout couple (P_0, P_1) qui minimiserait $\int \Phi\left(\frac{dR_0}{d(R_0 + R_1)}\right) d(R_0 + R_1)$ pour toutes les fonctions Φ de classe \mathcal{C}^2 constituerait

également une expérience minimale et serait donc équivalent (au sens de l'ordre sur les expériences) au couple de HUBER et STRASSEN. On va donner une condition suffisante beaucoup plus faible :

THÉORÈME (6.3) :

Sous les hypothèses du théorème (5.3), s'il existe une fonction Φ de classe \mathcal{C}^2 et strictement convexe, telle que (P_0, P_1) minimise la fonction

$$H(R_0, R_1) = \int \Phi \left(\frac{d R_0}{d(R_0 + R_1)} \right) d(R_0 + R_1)$$

sur $\mathcal{P}_{V_0} \times \mathcal{P}_{V_1}$, ce couple minimise toute fonction

$$K(R_0, R_1) = \int \Psi \left(\frac{d R_0}{d(R_0 + R_1)} \right) d(R_0 + R_1)$$

où Ψ est convexe de classe \mathcal{C}^2 . L'expérience constituée par $(E, \mathcal{T}; P_0, P_1)$ est donc minimale.

Preuve :

Quitte à ajouter des constantes à toutes les intégrales, ce qui ne change rien aux inégalités, on peut faire sur Φ et Ψ les mêmes hypothèses que pour le théorème (6.2).

On reprend les notations de ce dernier théorème, ainsi que la construction de (Q'_0, Q'_1) à partir du couple (Q_0, Q_1) du théorème (5.3).

Par hypothèse on a $I(1) = H(P_0, P_1) = H(Q_0, Q_1) \leq H(Q'_0, Q'_1) = I(0)$ ce qui implique que $I'(0)$ est négatif, I étant convexe.

Les inégalités (11) et (12) entraînent alors que l'on a

$$0 \leq \int (Q_1 - P_1) \{\varphi_1 \circ \gamma \geq t\} dt + \int (P_0 - Q_0) \{\psi_1 \circ \gamma \geq t\} dt \leq I'(0) + 4 \epsilon M \leq 4 \epsilon M$$

Comme ϵ est arbitraire et que ces intégrales sont positives par (12) on aura pour presque tout t

$$P_1 \{ \varphi_1 \circ \gamma \geq t \} = Q_1 \{ \varphi_1 \circ \gamma \geq t \} \quad P_0 \{ \psi_1 \circ \gamma \geq t \} = Q_0 \{ \psi_1 \circ \gamma \geq t \}$$

Φ étant strictement convexe, φ_1 et ψ_1 sont strictement décroissantes ce qui entraîne aussi pour presque tout t

$$(13) \quad P_1 \{ \gamma > t \} = Q_1 \{ \gamma > t \} \quad \text{et} \quad P_0 \{ \gamma > t \} = Q_0 \{ \gamma > t \}$$

Il reste à voir que l'on a $K(Q_0, Q_1) = K(P_0, P_1)$.

On notera

$$L(t) = K[(1-t)Q'_0 + tP_0, (1-t)Q'_1 + tP_1]$$

$$\varphi_2(z) = \frac{1}{1+z} \Psi\left(\frac{1}{1+z}\right) - \Psi\left(\frac{1}{1+z}\right)$$

$$\psi_2(z) = \frac{z}{1+z} \Psi\left(\frac{1}{1+z}\right) + \Psi\left(\frac{1}{1+z}\right)$$

φ_2 et ψ_2 ont les mêmes propriétés que φ_1 et ψ_1 et on a

$$L'(0) = \int (Q'_1 - P_1) \{ \varphi_2 \circ \gamma \geq t \} dt + \int (P_0 - Q'_0) \{ \psi_2 \circ \gamma \geq t \} dt$$

et donc d'après (13)

$$L'(0) \leq \int (Q_1 - P_1) \{ \varphi_2 \circ \gamma \geq t \} dt + \int (P_0 - Q_0) \{ \psi_2 \circ \gamma \geq t \} dt + 4 \in M' \leq 4 \in M'$$

où M' est un majorant commun à φ_2 et ψ_2 .

$$L''(t) = \int \Psi'' \left(\frac{(1-t)q'_0 + t p_0}{(1-t)(q'_0 + q'_1) + t(p_0 + p_1)} \right) \frac{(p_0 q'_1 - q'_0 p_1)^2}{[(1-t)(q'_0 + q'_1) + t(p_0 + p_1)]^3} d\mu$$

$$\leq D(t) \|\Psi''\|$$

si on pose
$$D(t) = \int \frac{(p_0 q'_1 - q'_0 p_1)^2}{[(1-t)(q'_0 + q'_1) + t(p_0 + p_1)]^3} d\mu$$

Φ étant strictement convexe, Φ'' admet un minimum $\delta > 0$ sur $[0, 1]$ et on peut donc écrire :

$$I''(t) \geq \delta D(t)$$

D'autre part on a vu que $I'(0)$ est supérieur à $-4 \in M$. Donc pour t dans $[0, 1[$

on peut écrire :

$$I(t) = I(0) + t I'(0) + \int_0^t I''(u)(t-u) du \geq I(0) + t I'(0) + \delta \int_0^t D(u)(t-u) du$$

$$\text{d'où } \int_0^t D(u)(t-u) du \leq I(t) - I(0) - t I'(0) \leq \frac{I(t) - I(0) + 4 \epsilon M t}{\delta}$$

et

$$L(t) = L(0) + t L'(0) + \int_0^t L''(u)(t-u) du \leq L(0) + 4 \epsilon M' + \|\Psi''\| \int_0^t D(u)(t-u) du$$

$$L(t) \leq L(0) + 4 \epsilon M' + \frac{\|\Psi''\|}{\delta} [I(t) - I(0) + 4 \epsilon M t]$$

Par continuité de L et I pour $t = 1$ on obtient

$$L(1) \leq L(0) + 4 \epsilon M' + \frac{\|\Psi''\|}{\delta} [I(1) - I(0) + 4 \epsilon M]$$

Comme $I(1) = H(P_0, P_1)$ est inférieur à $I(0) = H(Q'_0, Q'_1)$ on obtient facilement

$$K(P_0, P_1) = L(1) \leq K(Q'_0, Q'_1) + 4 \epsilon M' + 4 \epsilon M \frac{\|\Psi''\|}{\delta}$$

De plus on a $K(Q'_0, Q'_1) \leq K(Q_0, Q_1) + 4 \epsilon \|\Psi\|$.

ϵ étant arbitraire, ceci montre que (P_0, P_1) minimise K comme le couple (Q_0, Q_1) .

C.Q.F.D.

c) Cas des expériences ternaires :

On peut se demander si le résultat qui précède s'étend au cas d'expériences comportant plus de deux paramètres. Plus précisément, étant données $(n+1)$ capacités bi-alternatives v_0, \dots, v_n , les ensembles associés $\varphi_{v_0}, \dots, \varphi_{v_n}$ et les expériences de la forme $\mathcal{E} = \{E, \mathcal{A}; P_0, \dots, P_n\}$ où P_i est un élément de φ_{v_i} , existe-t-il une expérience minimale au sens de l'ordre sur les expériences ?

Le contre-exemple qui suit montre que ce résultat est faux dans le cas de trois capacités.

Contre-exemple :

L'espace E est l'ensemble $\{0 ; 1\}$ muni de la topologie discrète. Une capacité v sur E est définie par :

$$v(\{0\}) = \alpha \quad v(\{1\}) = \beta \quad v(\emptyset) = 0 \quad v(E) = 1$$

α et β étant compris entre 0 et 1.

Elle est bi-alternative dès que $v(E) \leq v(\{0\}) + v(\{1\})$ c'est-à-dire $\alpha \geq 1 - \beta$.

A une telle capacité est associé un ensemble φ_v formé des probabilités P sur E vérifiant $1 - \beta \leq P(\{0\}) \leq \alpha$. φ_v s'identifie donc à un intervalle $[1 - \beta, \alpha]$ compact de $[0, 1]$.

Pour $i = 0, 1, 2$ on définit des capacités v_i et les ensembles φ_{v_i} correspondants:

$$\varphi_{v_i} = \{x \delta_0 + (1 - x) \delta_1 \mid x \in [a_i, b_i] \quad 0 \leq a_0 < b_0 < a_1 < b_1 < a_2 < b_2 \leq 1\}$$

δ_t étant la masse de DIRAC en t.

S'il existe une expérience $\mathcal{E} = (E, \mathcal{A} ; P_0, P_1, P_2)$ minimale, d'après le théorème (6.1) de BLACKWELL, le triplet (P_0, P_1, P_2) minimise l'intégrale :

$$\int \Phi \left(\frac{d P_0}{d (P_0 + P_1 + P_2)}, \frac{d P_1}{d (P_0 + P_1 + P_2)}, \frac{d P_2}{d (P_0 + P_1 + P_2)} \right) d (P_0 + P_1 + P_2)$$

pour toute fonction Φ continue convexe sur le simplexe

$$S = \{x, y, z \geq 0 \mid x + y + z = 1\}$$

Pour la fonction $\Phi(x, y, z) = x^2$, on obtient un minimum unique avec le triplet

$$P_0(\{0\}) = b_0 \quad P_1(\{0\}) = a_1 \quad P_2(\{0\}) = a_2$$

Pour la fonction $\Psi(x, y, z) = z^2$ on obtient également un minimum unique au point

$$P_0(\{0\}) = b_0 \quad P_1(\{0\}) = b_1 \quad P_2(\{0\}) = a_2$$

différent du précédent. Ceci prouve qu'il ne peut exister dans ce cas d'expérience

minimale. Pour plus de précisions sur les expériences binaires et ternaires on pourra voir [91] .

Lucien BIRGÉ
Mathématiques
Université Paris VII
2, Place Jussieu
75221 PARIS CEDEX 05.

CHAPITRE VIII
EXEMPLES DE TESTS MINIMAX-ROBUSTES
INTERVALLES DE CONFIANCE ROBUSTES

par

Claude DENIAU, Georges OPPENHEIM, Claude VIANO

Nous aborderons en première partie quelques cas particuliers du problème abordé au Chapitre VII de la construction de tests robustes. En deuxième partie les résultats seront utilisés pour la construction d'intervalles de confiance robustes.

PREMIÈRE PARTIE : EXEMPLES DE TESTS MINIMAX DE DEUX HYPOTHÈSES
SIMPLES.

1. Modèles probabilistes et hypothèses.

Sur l'espace mesuré $(\mathbb{R}, \mathcal{B}, \mu)$, P_0 et P_1 sont deux mesures de probabilité absolument continues par rapport à μ . Le rapport $\frac{p_1(x)}{p_0(x)}$ des densités sera supposé (μ p.s.) monotone non décroissant. Nous reviendrons en 4 sur cette hypothèse. Le problème est de trouver un test minimax de P_0 contre P_1 , les déviations autour de P_0 et P_1 étant formalisées par deux domaines \mathcal{H}_0 et \mathcal{H}_1 disjoints de lois de probabilité sur $(\mathbb{R}, \mathcal{B})$.

Le modèle de base est $(\mathbb{R}^n, \mathcal{B}^{\otimes n}, P_i^n)$: les observations sont indépendantes et équidistribuées selon P_0 ou P_1 .

Le modèle étendu est $(\mathbb{R}^n, \mathcal{B}^{\otimes n}, \mathcal{H}_i^{\otimes n})$: ce modèle qui contient le modèle de base exclut toute dépendance entre les observations mais permet qu'elles ne soient pas équidistribuées. Les domaines \mathcal{H}_0 et \mathcal{H}_1 seront les suivants :

$$\mathcal{H}_0(\epsilon_0, \delta_0) = \{Q \mid Q\{x < t\} \geq (1 - \epsilon_0) P_0\{x < t\} - \delta_0 \quad \forall t\}$$

$$\mathcal{H}_1(\epsilon_1, \delta_1) = \{Q \mid Q\{x < t\} \leq (1 - \epsilon_1) P_1\{x < t\} + \epsilon_1 + \delta_1 \quad \forall t\}$$

(On notera que si $\frac{p_1}{p_0}$ est monotone \mathcal{H}_0 et \mathcal{H}_1 sont disjoints dès que $\epsilon_0 + 2\epsilon_1 + \delta_0 + \delta_1 < P_0(a) - P_1(a)$ où $a = \sup \{t \mid \frac{p_1}{p_0}(t) \leq 1\}$ puisqu'alors pour $t < a$, $(1 - \epsilon_0) P_0(t) - \delta_0 > (1 - \epsilon_1) P_1(t) + \epsilon_1 + \delta_1$). Ces domaines ont été introduits pour deux raisons, la première est que leur dissymétrie permettra le passage à la construction des intervalles de confiance, la deuxième est qu'ils permettent l'application des résultats à diverses perturbations classiques. En effet on a les inclusions suivantes :

a) Voisinage de P_i pour la distance de KOLMOGOROV :

$$K_i(\delta_i) = \{Q \mid \sup_t |Q\{X \leq t\} - P_i\{x \leq t\}| \leq \delta_i\} \subset \mathcal{H}_i(0, \delta_i)$$

b) Voisinage pour la distance en variation totale :

$$V_i(\delta_i) = \{Q \mid \sup_{B \in \mathcal{B}} |Q(B) - P_i(B)| \leq \delta_i\} \subset \mathcal{B}_i(0, \delta_i)$$

c) Modèle de contamination :

$$\mathcal{C}_i(\epsilon_i) = \{Q \mid Q = (1 - \epsilon_i) P_i + \epsilon_i H, H \in \mathcal{M}\} \subset \mathcal{B}_i(\epsilon_i, 0) .$$

2. Solution du problème du minimax.

Théorème 1 (HUBER [48 bis]) .

Sous les hypothèses du paragraphe 1, soient les deux fonctions q_0 et q_1 définies par :

$$\text{Si } \frac{p_1(x)}{p_0(x)} < K_0, \quad q_0(x) = \frac{1 - \epsilon_0}{v_1 + w_0 K_0} [v_1 p_0(x) + w_0 p_1(x)] \quad \text{et}$$

$$q_1(x) = \frac{(1 - \epsilon_1) K_0}{v_1 + w_0 K_0} [v_1 p_0(x) + w_0 p_1(x)]$$

$$(1) \quad \text{Si } K_0 < \frac{p_1(x)}{p_0(x)} < K_1, \quad q_0(x) = (1 - \epsilon_0) p_0(x) \quad \text{et} \quad q_1(x) = (1 - \epsilon_1) p_1(x)$$

$$\text{Si } \frac{p_1(x)}{p_0(x)} > K_1, \quad q_0(x) = \frac{1 - \epsilon_0}{w_1 + v_0 K_1} [w_1 p_0(x) + v_0 p_1(x)] \quad \text{et}$$

$$q_1(x) = \frac{(1 - \epsilon_1) K_1}{w_1 + v_0 K_1} [w_1 p_0(x) + v_0 p_1(x)]$$

$$\text{avec } v_j = (\epsilon_j + \delta_j)/(1 - \epsilon_j) > 0 \quad w_j = \delta_j/(1 - \epsilon_j) < 1 \quad (j = 0, 1) \quad ,$$

K_0 et K_1 étant solutions de

$$(2) \quad \begin{cases} (K_0 P_0 \{ \frac{p_1}{p_0} < K_0 \} - P_1 \{ \frac{p_1}{p_0} < K_0 \}) = v_1 w_0 K_0 \\ (P_1 \{ \frac{p_1}{p_0} > K_1 \} - K_1 P_0 \{ \frac{p_1}{p_0} > K_1 \}) = v_0 K_1 + w_1 \end{cases}$$

a) $q_0(x)$ et $q_1(x)$ sont les densités de deux lois de probabilité

$$Q_0 \in \mathcal{B}_0 \quad \text{et} \quad Q_1 \in \mathcal{B}_1 .$$

b) le couple Q_0, Q_1 est minimal (Chapitre VII, définition 0.2) pour le test de P_0 contre P_1 avec les perturbations \mathcal{B}_0 et \mathcal{B}_1 .

c) pour tout autre couple minimal Q'_0, Q'_1 le rapport $\frac{d Q'_1}{d Q'_0}$ est ($P_0 + P_1$ p.s.) égal à $\frac{d Q_1}{d Q_0}$.

Preuve :

a) On montre d'abord l'existence et l'unicité de K_0 et K_1 ainsi que $K_0 < K_1$:

Considérons

$$f(c) = \frac{c P_0 \left\{ \frac{p_1}{p_0} < c \right\} - P_1 \left\{ \frac{p_1}{p_0} < c \right\}}{v_1 + w_0 c} \quad \text{et} \quad g(c) = \frac{P_1 \left\{ \frac{p_0}{p_1} < \frac{1}{c} \right\} - c P_0 \left\{ \frac{p_0}{p_1} < \frac{1}{c} \right\}}{v_0 c + w_1}$$

Des calculs assez simples montrent que f et g sont continues. f est nulle si $c \leq \text{ess inf } \frac{p_1}{p_0}$. Elle est strictement croissante dès qu'elle est non nulle et tend vers $\frac{1}{w_0}$ lorsque c tend vers l'infini. g est nulle si $c \geq \text{ess sup } \frac{p_1}{p_0}$. Elle est strictement décroissante sinon et tend vers $\frac{1}{w_1}$ lorsque c tend vers zéro. Donc K_0 et K_1 existent et sont uniques. Enfin $f(1) > 1$ et $g(1) > 1$ et donc $K_0 < 1 < K_1$ dès que

$$v_1 + w_0 \leq P_0 \left\{ \frac{p_1}{p_0} < 1 \right\} - P_1 \left\{ \frac{p_1}{p_0} < 1 \right\} \quad (\neq 0)$$

$$\text{et} \quad v_0 + w_1 < P_1 \left\{ \frac{p_0}{p_1} < 1 \right\} - P_0 \left\{ \frac{p_0}{p_1} < 1 \right\} \quad (\neq 0)$$

En ce qui concerne les fonctions de répartition on a par exemple

$$\begin{aligned} \text{si} \quad & \frac{p_1(x)}{p_0(x)} \leq K_0 \quad (1 - \epsilon_0) P_0(x) - \delta_0 \leq Q_0(x) \leq (1 - \epsilon_0) P_0(x) \\ (3) \quad \text{si} \quad & K_0 < \frac{p_1(x)}{p_0(x)} < K_1 \quad Q_0(x) = (1 - \epsilon_0) P_0(x) - \delta_0 \\ \text{si} \quad & \frac{p_1(x)}{p_0(x)} \geq K_1 \quad (1 - \epsilon_0) P_0(x) - \delta_0 + \epsilon_0 \leq Q_0(x) \leq (1 - \epsilon_0) P_0(x) + \epsilon_0 \end{aligned}$$

ce qui prouve que $Q_0 \in \mathcal{B}_0$ (on montre aisément que q_0 est une densité)

b) pour prouver que le couple Q_0, Q_1 est minimal il suffit de montrer (Chapitre VII, proposition 1.1) que quelles que soient $Q \in \mathcal{B}_0$ et $Q' \in \mathcal{B}_1$ et pour tout a

$$(4) \quad Q \left\{ \frac{q_1(x)}{q_0(x)} < a \right\} \geq Q_0 \left\{ \frac{q_1(x)}{q_0(x)} < a \right\} \geq Q_1 \left\{ \frac{q_1(x)}{q_0(x)} < a \right\} \geq Q' \left\{ \frac{q_1(x)}{q_0(x)} < a \right\} .$$

Or $\frac{q_1(x)}{q_0(x)}$ est proportionnel au rapport tronqué $\left[\frac{p_1(x)}{p_0(x)} \right]_{K_0}^{K_1}$: il s'ensuit que pour tout a

$$Q \left\{ \frac{q_1(x)}{q_0(x)} < a \right\} \geq (1 - \epsilon_0) P_0 \left\{ \frac{q_1(x)}{q_0(x)} < a \right\} - \delta_0 = Q_0 \left\{ \frac{q_1(x)}{q_0(x)} < a \right\}$$

(trivialement si $a \notin [K_0, K_1]$ et comme conséquence de (3) sinon) . La dernière inégalité se prouve de la même manière.

c) l'unicité du rapport $\frac{dQ_1}{dQ_0}$ découle des théorèmes 5.3 et 5.5 du Chapitre VII .

Remarques concernant le couple minimal :

$$R_1 : \text{Posons } k_0 = \inf \left\{ x \mid \frac{p_1(x)}{p_0(x)} \geq K_0 \right\} \text{ et } k_1 = \sup \left\{ x \mid \frac{p_1(x)}{p_0(x)} \leq K_1 \right\} .$$

q_0 est obtenue en enlevant à la densité sous stochastique $(1 - \epsilon_0) p_0$ la masse δ_0 à gauche de k_0 et en lui ajoutant la masse $\delta_0 + \epsilon_0$ à droite de k_1 , la fonction de répartition Q_0 étant à la frontière de \mathcal{H}_0 entre k_0 et k_1 . Le phénomène est analogue pour Q_1 , la masse transportée étant enlevée à droite de k_1 : la fonction de répartition Q_1 étant égale à $(1 - \epsilon_1) P_1 + \epsilon_1 + \delta_1$ entre k_0 et k_1 .

R_2 : le couple Q_0, Q_1 n'est pas unique ; il est clair que l'on pourrait par exemple conserver q_0 et q_1 sur le segment $[k_0, k_1]$, et les modifier à l'extérieur à condition d'y conserver la masse totale et le rapport $\frac{q_1(x)}{q_0(x)}$.

R_3 : le rapport des densités $\frac{q_1(x)}{q_0(x)} = \frac{1 - \epsilon_1}{1 - \epsilon_0} \left[\frac{p_1(x)}{p_0(x)} \right]_{K_0}^{K_1}$ est proportionnel au rapport tronqué des densités des lois P_0 et P_1 .

Il résulte de ce qui précède que, pour tout test du rapport de vraisemblance entre Q_0 et Q_1 , sa pire performance sur $\mathcal{H}_0^{\otimes n}$ et $\mathcal{H}_1^{\otimes n}$ est atteinte en Q_0^n, Q_1^n :

Théorème 2 :

La famille des tests du rapport de vraisemblance entre Q_0 et Q_1

$$\sum_i \left[\frac{p_1(x_i)}{p_0(x_i)} \right] \frac{K_1}{K_0} > K \Rightarrow \varphi(x) = 1 \quad [\text{où } x = (x_1, x_2, \dots, x_n)]$$

$$\sum_i \left[\frac{p_1(x_i)}{p_0(x_i)} \right] \frac{K_1}{K_0} < K \Rightarrow \varphi(x) = 0$$

constitue une famille de tests minimax pour tester P_0 contre P_1 avec les perturbations δ_0 et δ_1 .

(La démonstration a été faite au Chapitre VII, théorème 1.2).

3. Application à quelques exemples classiques de modèles étendus.

En appliquant les résultats précédents aux domaines $\delta_i(0, \delta_i)$ et $\delta_i(\epsilon_i, 0)$ on obtient une famille de tests minimax de P_0 contre P_1 lorsque les perturbations sont des boules pour la distance en variation totale ou lorsqu'elles sont de type "contamination".

Théorème 3 :

Sous les hypothèses des théorèmes 1 et 2, pour le test de P_0 contre P_1 , les perturbations autour de P_i ($i = 0, 1$) étant

$$V_i(\delta_i) = \{Q \mid \sup_{B \in \mathcal{B}} |Q(B) - P_i(B)| \leq \delta_i\}$$

a) un couple minimal est donné par

$$\frac{p_1(x)}{p_0(x)} \leq K_0 : q_0(x) = \frac{\delta_1 p_0(x) + \delta_0 p_1(x)}{\delta_1 + \delta_0 K_0} \quad \text{et} \quad q_1(x) = \frac{\delta_1 p_0(x) + \delta_0 p_1(x)}{\delta_1 + \delta_0 K_1}$$

$$(5) \quad K_0 < \frac{p_1(x)}{p_0(x)} < K_1 : q_0(x) = p_0(x) \quad \text{et} \quad q_1(x) = p_1(x)$$

$$\frac{p_1(x)}{p_0(x)} \geq K_1 : q_0(x) = K_0 \frac{\delta_1 p_0(x) + \delta_0 p_1(x)}{\delta_1 + \delta_0 K_0} \quad \text{et} \quad q_1(x) = K_1 \frac{\delta_1 p_0(x) + \delta_0 p_1(x)}{\delta_1 + \delta_0 K_1}$$

$$(6) \quad \text{avec} \quad \begin{cases} K_0 P_0 \left\{ \frac{p_1}{p_0} < K_0 \right\} - P_1 \left\{ \frac{p_1}{p_0} < K_0 \right\} = \delta_1 + \delta_0 K_0 \\ P_1 \left\{ \frac{p_1}{p_0} > K_1 \right\} - K_1 P_0 \left\{ \frac{p_1}{p_0} > K_1 \right\} = \delta_0 K_1 + \delta_1 \end{cases}$$

b) La famille des tests du rapport de vraisemblance tronqué

$$\sum_i \left[\frac{p_1(x_i)}{p_0(x_i)} \right]^{K_1} > K \Rightarrow \varphi(x) = 1 \quad [\text{où } x = (x_1, \dots, x_n)]$$

$$\sum_i \left[\frac{p_1(x_i)}{p_0(x_i)} \right]^{K_1} < K \Rightarrow \varphi(x) = 0$$

constitue une famille minimax.

Preuve :

Il suffit d'utiliser l'inclusion $V_i(\delta_i) \subset \mathcal{H}_i(0, \delta_i)$ et de montrer que les lois Q_0 et Q_1 appartiennent bien aux domaines V_0 et V_1 : ainsi

$$\forall A \in \mathcal{B}: Q_0(A) - P_0(A) = (Q_0 - P_0)(A \cap \{ \frac{p_1}{p_0} < K_0 \}) + (Q_0 - P_0)(A \cap \{ \frac{p_1}{p_0} > K_1 \})$$

le premier terme est négatif et supérieur à $-\delta_0$ et le second positif et inférieur à δ_0 d'après (3), donc $Q_0 \in V_0(\delta_0)$.

Théorème 3' :

Pour le test de P_0 contre P_1 , les perturbations étant

$$C_i(\epsilon_i) = \{Q = (1 - \epsilon_i) P_i + \epsilon_i H, H \in \mathcal{H}\} \quad (i = 0, 1)$$

La famille des tests du rapport de vraisemblance tronqué en K_0 et K_1 est minimax, le couple minimal étant donné par :

$$(7) \quad \begin{aligned} q_0(x) &= (1 - \epsilon_0) p_0(x) & \text{si } \frac{p_1(x)}{p_0(x)} \leq K_1 \\ q_0(x) &= \frac{(1 - \epsilon_0)}{K_1} p_1(x) & \text{si } \frac{p_1(x)}{p_0(x)} > K_1 \\ q_1(x) &= (1 - \epsilon_1) K_0 p_0(x) & \text{si } \frac{p_1(x)}{p_0(x)} \leq K_0 \\ q_1(x) &= (1 - \epsilon_1) p_1(x) & \text{si } \frac{p_1(x)}{p_0(x)} > K_0 \end{aligned}$$

avec
$$K_0 P_0 \left\{ \frac{p_1}{p_0} < K_0 \right\} - P_1 \left\{ \frac{p_1}{p_0} < K_0 \right\} = \epsilon_1 / 1 - \epsilon_1$$

$$P_1 \left\{ \frac{p_1}{p_0} > K_1 \right\} - K_1 P_0 \left\{ \frac{p_1}{p_0} > K_1 \right\} = \frac{\epsilon_0 K_1}{1 - \epsilon_0}$$

Preuve :

Etant donnée l'inclusion $\mathcal{C}_1(\epsilon_1) \subset \mathcal{B}_1(\epsilon_1, 0)$ il suffit de remarquer que Q_0 et Q_1 appartiennent bien à \mathcal{C}_0 et \mathcal{C}_1 : ainsi Q_0 est obtenue à partir de P_0 en la contaminant par la loi H de support $[k_1, +\infty[$ et de densité $\frac{1 - \epsilon_0}{\epsilon_0} \left(\frac{p_1(x)}{K_1} - p_0(x) \right)$.

4. Cas où le rapport des densités n'est pas monotone.

Les résultats qui suivent peuvent se démontrer directement (HUBER [45]). Nous les faisons apparaître ici comme conséquence des théorèmes valides lorsque le rapport des densités est monotone.

Sur $(\Omega, \mathcal{A}, \mu)$, P_0 et P_1 sont deux lois de probabilité de densités p_0 et p_1 par rapport à μ .

Théorème 4 :

Le couple (Q_0, Q_1) dont les densités ont été données en (5) est minimal pour tester P_0 contre P_1 lorsque les domaines de perturbation sont $V_0(\delta)$ et $V_1(\delta)$, boules de même rayon pour la distance ou variation totale.

Preuve :

Notons d'abord (la démonstration faite au théorème 3 reste valable) que $Q_1 \in V_1(\delta)$. Il reste alors à montrer que :

a) les équations (6) ont une solution unique $K_0 < K_1$

b) Q_0 et Q_1 satisfont aux trois inégalités (4).

Pour cela, posant $T(\omega) = \frac{p_1(\omega)}{p_0(\omega)}$, on transporte $(\Omega, \mathcal{A}, \mu)$ en $(\mathbb{R}^+, \mathcal{B}, T\mu)$: On notera ${}^T P_0$ et ${}^T P_1$ les images de P_0 et P_1 , V_0^* et V_1^* les boules de rayon δ autour de ${}^T P_0$ et ${}^T P_1$.

${}^T P_0$ et ${}^T P_1$ sont absolument continues par rapport à $T\mu$ et, si g_0 et g_1 sont leurs densités, $\frac{g_1(x)}{g_0(x)} = x$ ($T\mu$ p.s.). En effet : (en convenant que $g_0 = p_0 = 0 \Rightarrow \frac{g_1}{g_0} = T$ et en constatant que, μ p.s., $g_0 = 0 \Leftrightarrow p_0 = 0$)

$$\begin{aligned} \int_{g_1 > T g_0} (g_1(T(\omega)) - T(\omega) g_0(T(\omega))) \mu(d\omega) &= \int_{g_1 > T g_0} p_1(\omega) \mu(d\omega) \\ &- \int_{g_1 > T g_0} T(\omega) p_0(\omega) \mu(d\omega) = 0 \end{aligned}$$

entraîne $\mu\{g_1 > T g_0\} = 0$ et de même $\mu\{g_1 < T g_0\} = 0$.

. Le rapport $\frac{g_1}{g_0}(x) = x$ étant (T_μ p.s.) monotone croissant, les résultats du paragraphe 3 s'appliquent au test de T_{P_0} contre T_{P_1} , les domaines étant $V_0^*(\delta)$ et $V_1^*(\delta)$: On remarque immédiatement que le couple minimal fourni par le théorème 1 est (T_{Q_0}, T_{Q_1}) . Ceci prouve que :

i) les équations (6) ont une solution unique $K_0 < K_1$ si δ est suffisamment petit puisque

$$P_0\left\{\frac{p_1}{p_0} < K_0\right\} = T_{P_0}\{x < K_0\} \text{ et } P_1\left\{\frac{p_1}{p_0} < K_0\right\} = T_{P_1}\{x < K_0\}$$

ii) $\forall Q^* \in V_0^*(\delta), \forall Q'^* \in V_1^*(\delta),$

$$Q^*\left\{\left[\frac{p_1}{p_0}\right]_{K_0}^{K_1} \geq a\right\} \leq T_{Q_0}\left\{\left[x\right]_{K_0}^{K_1} \geq a\right\} \leq T_{Q_1}\left\{\left[x\right]_{K_0}^{K_1} \geq a\right\} \leq Q'^*\left\{\left[x\right]_{K_0}^{K_1} \geq a\right\}$$

Et on termine la démonstration de b) en remarquant que

$$Q \in V_0(\delta) \Rightarrow |Q\{T(\omega) < t\} - P_0\{T(\omega) < t\}| < \delta, \forall t \Rightarrow T_Q \in V_0^*(\delta)$$

et $Q' \in V_1(\delta) \Rightarrow T_{Q'} \in V_1^*(\delta)$, ce qui entraîne

$$Q\left\{\left[\frac{p_1}{p_0}\right]_{K_0}^{K_1} \geq a\right\} = T_Q\left\{\left[x\right]_{K_0}^{K_1} \geq a\right\} \leq T_{Q_0}\left\{\left[x\right]_{K_0}^{K_1} \geq a\right\} = Q_0\left\{\left[\frac{p_1}{p_0}\right]_{K_0}^{K_1} \geq a\right\} \text{ pour}$$

toute loi Q de $V_0(\delta)$.

Théorème 4' :

Le couple (Q_0, Q_1) dont les densités ont été données en (7) est minimal pour tester P_0 contre P_1 lorsque les domaines de perturbation sont $\mathcal{C}_0(\epsilon_0)$ et $\mathcal{C}_1(\epsilon_1)$.

La preuve, semblable à la précédente, ne sera pas répétée : elle est basée sur le fait que le couple (T_{Q_0}, T_{Q_1}) est minimal pour tester T_{P_0} contre T_{P_1} avec les domaines \mathcal{C}_0^* et \mathcal{C}_1^* et que, pour toute loi Q de $\mathcal{C}_1(\epsilon_1)$ sont image, $T_Q = (1 - \epsilon_1)T_{P_1} + \epsilon_1 T_H$, appartient à \mathcal{C}_1^* .

DEUXIÈME PARTIE : INTERVALLES DE CONFIANCE MINIMAX-ROBUSTE.

Les résultats précédents vont être utilisés pour construire une famille d'intervalles de confiance d'un paramètre de translation (HUBER[48 bis]). Cette famille aura des propriétés de minimax-robustesse analogues à celles des tests dont elle est issue.

1. Le modèle probabiliste, les hypothèses et les notations.

Soit, sur \mathbb{R} , \mathcal{B} muni de la mesure de LEBESGUE λ une fonction de répartition F telle que :

- . F est absolument continue par rapport à λ .
- . Sa densité f est continue et $-\log f$ est strictement convexe.

On considérera la famille $F_\theta(x) = F(x - \theta)$ paramétrée par θ ($\theta \in \mathbb{R}$). Le problème est de bâtir un intervalle de confiance pour θ en tenant compte d'un domaine \mathcal{B} de perturbation autour de F :

Le modèle de base est $\{\mathbb{R}^n, \mathcal{B}^{\otimes n}, F_\theta^n \mid \theta \in \mathbb{R}\}$.

Le modèle étendu est $\{\mathbb{R}^n, \mathcal{B}^{\otimes n}, (F^\theta)^n \mid \theta \in \mathbb{R}\}$

où F^θ est le domaine translaté de \mathcal{B} : $F^\theta = \{G \mid G_{-\theta} \in \mathcal{B}\}$.

Par la suite on désignera par P^θ la loi du domaine F^θ correspondant à la loi P de \mathcal{B} .

Dans \mathbb{R}^n on appellera translation et on notera $x \rightarrow x + k$ la transformation $(x_1, x_2, \dots, x_n) \rightarrow (x_1 + k, x_2 + k, \dots, x_n + k)$.

Enfin $(P^\theta)^n$ sera notée P^θ lorsqu'il n'en résultera pas d'ambiguïté.

2. Le minimax.

a est un nombre positif fixé. A tout estimateur T , application $\mathcal{B}^{\otimes n}$ mesurable de \mathbb{R}^n dans \mathbb{R} , on attache la quantité

$$S(T) = \sup \{ \text{Max} (P^\theta \{T < \theta - a\}, P^\theta \{T > \theta + a\}) \mid \theta \in \mathbb{R}, P \in \mathcal{B}^n \}$$

où $\mathcal{B} = \mathcal{B}_0(\epsilon_0, \delta_0) \cap \mathcal{B}_1(\epsilon_1, \delta_1)$, domaines définis en première partie.

Définition : T^0 sera dit minimax robuste si il minimise $S(T)$:

$$\forall T : S(T^0) \leq S(T)$$

Quel que soit n , l'intervalle bilatéral de longueur fixe $[T^0 - a, T^0 + a]$ aura un coefficient de confiance supérieur à $1 - 2S(T^0)$ et les deux intervalles unilatéraux $[T^0 - a, +\infty[$ et $]-\infty, T^0 + a]$ un coefficient de confiance supérieur à $1 - S(T^0)$.

3. Construction de T^0 .

D'après les hypothèses faites sur F , la famille F_θ est à rapport de densités strictement croissant. Il en résulte, par application du Théorème 2 que le test

$$h(x) < K \Leftrightarrow \varphi(x) = 0$$

$$h(x) > K \Leftrightarrow \varphi(x) = 1 \quad \text{où } h(x) = \sum_1^n \log \left[\frac{f(x_i - a)}{f(x_i + a)} \right] \frac{K_1}{K_0}$$

$$h(x) = K \Leftrightarrow \varphi(x) = \chi$$

est minimax pour tester F_{-a} contre F_a pour les domaines $\mathcal{B}_0^{-a}(\epsilon_0, \delta_0)$ et $\mathcal{B}_1^a(\epsilon_1, \delta_1)$ entourant ces lois. K et χ seront choisis pour que φ minimise $\sup \{ \text{Max}(E_Q(\varphi), E_{Q'}(1 - \varphi)) \mid Q \in \mathcal{B}_0^{-a}, Q' \in \mathcal{B}_1^a \}$. On pose alors :

$\alpha = E_{Q_0^{-a}}(\varphi) = E_{Q_1^a}(1 - \varphi)$, (Q_0^{-a}, Q_1^a) étant le couple minimal fourni par le théorème 1.

Alors $\varphi(x - \theta)$ fournit un test de $F_{\theta-a}$ contre $F_{\theta+a}$ ayant les mêmes propriétés que le précédent pour les domaines translatés. Le passage aux intervalles de confiance s'effectue alors de la façon habituelle en posant :

$$T^*(x) = \inf \{ \theta \mid h(x - \theta) \leq K \}$$

$$T^{**}(x) = \sup \{ \theta \mid h(x - \theta) \geq K \}$$

$(T^*$ et T^{**} sont finis et $T^* \leq T^{**}$ puisque $h(x - \theta)$ est décroissante en θ , qu'elle tend vers $n \log K_0$ et $n \log K_1$ lorsque θ tend vers l'infini et que $n \log K_0 < K < n \log K_1$ car $0 < \alpha \leq \frac{1}{2}$).

Notons $T^0(x)$ l'estimateur randomisé prenant les valeurs $T^*(x)$ et $T^{**}(x)$ avec les probabilités $1 - \chi$ et χ .

T^0 a les propriétés suivantes :

. Il est, comme T^* et T^{**} , équivariant par translation.

. Quel que soit θ et pour toute loi P il vérifie les inégalités

$$(8) \quad P \{T^0 < \theta\} \leq E_P (1 - \varphi(x - \theta)) \quad \text{et} \quad P \{T^0 > \theta\} \leq E_P (\varphi(x - \theta))$$

L'égalité étant atteinte dès que $P \{T^* = \theta\} = P \{T^{**} = \theta\} = 0$

En effet,

$$\begin{aligned} P \{T^0 < \theta\} &= (1 - \chi) P \{T^* < \theta\} + \chi P \{T^{**} < \theta\} \\ &\leq (1 - \chi) P \{T^* \leq \theta\} + \chi P \{T^{**} < \theta\} \\ &= (1 - \chi) P \{h(x - \theta) \leq K\} + \chi P \{h(x - \theta) < K\} = E_P (1 - \varphi(x - \theta)). \end{aligned}$$

Il reste à examiner la robustesse de T^0 :

Théorème 5 :

T^0 est minimax-robuste.

On montrera successivement que $S(T^0) = \alpha$ et que $S(T) \geq \alpha$ pour tout T . Le résultat suivant sera nécessaire :

On note \mathcal{C} l'ensemble des estimateurs presque équivariants par translation pour la famille de lois \mathfrak{F} :

$$T \in \mathcal{C} \Leftrightarrow \forall k \in \mathbb{R} \quad \forall P \in \mathfrak{F} \quad \forall x \in \mathbb{R}^n \quad P \{T(x+k) \neq T(x) + k\} = 0.$$

Lemme 1 :

Si \mathfrak{F} est stable par translation, $P \{T(x) = k\} = 0$ quel que soit k , pour tout T de \mathcal{C} et toute loi P de \mathfrak{F} . En effet, quels que soient $a, b \leq a$ et $\epsilon \in]0, a-b[$ on peut écrire

$$P \{b < T(x) < a\} + P \{a \leq T(x) < a + \epsilon\} = P \{b < T(x) \leq b + \epsilon\} + P \{b + \epsilon < T(x) < a + \epsilon\}$$

$$\text{Or } P \{b + \epsilon < T(x) < a + \epsilon\} = P^{-\epsilon} \{b + \epsilon < T(x + \epsilon) < a + \epsilon\}$$

$$\text{et, si } P \in \mathcal{F} \text{ et } T \in \mathcal{C}, \quad P^{-\epsilon} \{b + \epsilon < T(x + \epsilon) < a + \epsilon\} = P^{-\epsilon} \{b < T(x) < a\}.$$

Il s'ensuit que

$$P \{b < T(x) < a\} + P \{a \leq T(x) < a + \epsilon\} = P \{b < T(x) < b + \epsilon\} + P^{-\epsilon} \{b < T(x) < a\}$$

$$\text{Puis, en faisant tendre } \epsilon \text{ vers zéro : } P \{T(x) = a\} = 0.$$

Ce lemme sera utilisé par la suite avec la famille des translatées des deux éléments du couple minimal : $\mathcal{F} = \{Q_0^\theta \mid \theta \in \mathbb{R}\} \cup \{Q_1^\theta \mid \theta \in \mathbb{R}\}.$

i) Montrons maintenant que $S(T^0) = \alpha$. D'après les propriétés du couple minimal et la définition de α :

$$\forall P \in \mathcal{P} : P^{-a} \in \mathcal{P}_0^{-a}, \quad P^a \in \mathcal{P}_a, \text{ ce qui entraîne}$$

$$\text{Max} [E_{P^{-a}}(\varphi(x)), E_{P^a}(1 - \varphi(x))] \leq \text{Max} [E_{Q_0^{-a}}(\varphi(x)), E_{Q_1^a}(1 - \varphi(x))] = \alpha$$

Puis, en utilisant la propriété (8) et l'équivariance de T^0

$$\begin{aligned} \forall P \in \mathcal{P}, \forall \theta : \text{Max} [P^\theta \{T^0 < \theta - a\}, P^\theta \{T^0 > \theta + a\}] \\ \leq \text{Max} [E_{Q_1^a}(1 - \varphi(x)), E_{Q_0^{-a}}(\varphi(x))] = \alpha \end{aligned}$$

ceci prouve que $S(T^0) \leq \alpha$. L'égalité vient par application du lemme 1 à la propriété (8) :

$$\begin{aligned} E_{Q_0^{-a}}(\varphi(x)) &= Q_0^\theta \{T^0 > \theta + a\} \\ E_{Q_1^a}(1 - \varphi(x)) &= Q_1^\theta \{T^0 < \theta - a\} \end{aligned}$$

$$\text{soit } S(T^0) = \text{Max} (Q_1^\theta \{T^0 < \theta - a\}, Q_0^\theta \{T^0 > \theta + a\}) = \alpha$$

ii) Montrons ensuite que, pour tout estimateur presque équivariant par translation pour la famille \mathcal{F} , $S(T) \geq \alpha$:

Par application du lemme 1, $P \{T = 0\} = 0$ pour tout élément P de \mathcal{F} . On utilise alors $\{T < 0\}$ pour tester F_{-a} contre F_a . Par définition du couple minimal,

$\text{Max} [Q_0^{-a} \{T > 0\}, Q_1^a \{T < 0\}] \geq \alpha$ ce qui implique $S(T) \geq \alpha$.

iii) Il reste à prouver que $S(T) \geq \alpha$ pour un estimateur quelconque. Ceci se fait en utilisant le théorème de HUNT et STEIN que nous rappelons ici sous forme de Lemme :

Lemme 2 : (Théorème de HUNT et STEIN)

Sur $(\mathbb{R}^n, \mathcal{B}^n)$ une famille $\vartheta = \{P_\theta, \theta \in \Theta\}$ de distributions de probabilités, un groupe G de transformations sur \mathbb{R}^n et \mathcal{G} une σ -algèbre de parties de G , D l'ensemble des estimateurs et $R(\theta, d)$ une fonction de risque. Si

- a) ϑ est équivalente à une mesure μ σ -finie
- b) il existe une suite ν_n de distributions de probabilités sur (G, \mathcal{G}) telle que

$$\forall g \in G, \quad \forall c \in \mathcal{G}, \quad \lim_{n \rightarrow \infty} |\nu_n(cg) - \nu_n(c)| = 0.$$

Alors la classe des estimateurs presque équivariants pour G et ϑ est essentiellement complète (ZACKS [10]).

On applique ce lemme à la famille \mathcal{F} (la condition i) est vérifiée puisque \mathcal{F} est éliminée par la mesure de LEBESGUE. Pour la vérification de ii) on peut se reporter par exemple à LEHMANN [68]) et on prend comme risque

$$R(T, Q_0^\theta) = Q_0^\theta \{T > \theta + a\}, \quad R(T, Q_1^\theta) = Q_1^\theta \{T < \theta - a\}.$$

Quel que soit T il existe donc $T' \in \mathcal{C}$ tel que

$$\begin{aligned} Q_0^\theta \{T > \theta + a\} &\geq Q_0^\theta \{T' > \theta + a\} \\ \forall \theta & \\ Q_1^\theta \{T < \theta - a\} &\geq Q_1^\theta \{T' < \theta - a\} \end{aligned}$$

ce qui implique $S(T) \geq \alpha$ par confrontation avec les résultats de b). Le théorème est démontré.

4. Remarques et cas particulier de la loi de GAUSS.

- a) Si F est une distribution symétrique et avec $\delta_1 = \delta$, $\epsilon_1 = 0$, on peut prendre

0 et $\frac{1}{2}$ comme valeurs de K et χ , et $-K_1 = K_0 = -k$. T^* et T^{**} sont alors les extrémités de l'intervalle des solutions de l'équation

$$\sum_{i=1}^n \log \left[\frac{f(x_i - a - t)}{f(x_i + a - t)} \right]_{-k}^k = 0$$

et T^0 est l'estimateur randomisé prenant la valeur T^* avec la probabilité $\frac{1}{2}$.

b) En particulier si $F = \Phi$ (loi de GAUSS) l'équation devient :

$$\sum_{i=1}^n \Psi(x_i - t) = 0 \quad \text{avec} \quad \Psi(x) = \begin{cases} -2ak & \text{si } x < -k \\ 2ax & \text{si } -k \leq x \leq k \\ 2ak & \text{si } x > k \end{cases}$$

k , a et δ étant reliés par la relation

$$e^{-ak} \Phi(a - k) - e^{ak} \Phi(-a - k) = \delta (e^{ak} + e^{-ak})$$

c) Remarquons que $\frac{T^* + T^{**}}{2}$ est le M estimateur associé à la fonction

$$\psi(x) = \log \left[\frac{f(x - a)}{f(x + a)} \right]_{-k}^k.$$

Il s'ensuit que, si les conditions de convergence presque sûre de T^* et T^{**} sont réalisées (voir Chapitre IV), T^0 converge en probabilité vers leur limite commune.

C'est le cas en particulier pour la fonction ψ donnée en b) pour toute loi G symétrique telle que $E_G(\psi(x - t))$ ne s'annule que pour $t = 0$: T^0 converge alors en probabilité vers la médiane θ de G_θ .

d) La conséquence est que, dans les cas les plus réguliers, on est en présence d'une suite $[T_n^0 - a, T_n^0 + a]$ d'intervalles de confiance qui ont l'avantage d'être minimax-robustes quel que soit n , mais le double inconvénient

- . d'avoir un coefficient de confiance rarement calculable pour n fini,
- . d'avoir une amplitude a fixée alors qu'ils sont basés sur un estimateur convergent de θ .

L'idée vient alors immédiatement de considérer la suite T_n^{-1} d'estimateurs obtenus en faisant décroître a lorsque n croît. Voir à ce sujet le travail de C. HUBER-CAROL [54] qui a obtenu en faisant décroître a comme $\frac{1}{\sqrt{n}}$ et avec des hypothèses supplémentaires sur F , une minoration asymptotique du niveau de la suite des intervalles de confiance basés sur T_n^{-1} .

Claude DENIAU
Georges OPPENHEIM
Claude VIANO
UER de Mathématiques
Université Paris V
12, rue Cujas,
75005 PARIS.

CHAPITRE IX

ROBUSTESSE : π -ROBUSTESSE ET MINIMAX-ROBUSTESSE

par

Claude DENIAU - Georges OPPENHEIM - Claude VIANO.

Plusieurs critères de robustesse d'estimateurs ont été développés dans la pratique statistique. Nous abordons dans ce texte deux approches basées sur l'idée selon laquelle un estimateur est robuste si :

"ses performances sont peu modifiées par de faibles modifications de la loi modèle des observations". Pour rendre précise la proposition précédente il reste à définir deux points au moins :

- les performances étudiées,
- le type d'écart au modèle de base que l'on retient pour l'étude. Nous appelons dans ce qui suit contre-modèle le modèle spécifiant les écarts retenus.

Dans la définition 1 de la robustesse, celle de la minimax-robustesse, on utilise comme mesure de la performance les variances asymptotiques des estimateurs qui sont (le plus souvent) convergents en moyenne quadratique et asymptotiquement gaussiens lorsque centrés réduits.

Dans la définition 2, celle de la π -robustesse la performance est la valeur de la distance de PROHOROV entre les lois de l'estimateur calculées sous le modèle de base et sous le contre-modèle.

Considérant que la définition minimax possède un support intuitif traditionnel pour les statisticiens, nous consacrons la majeure partie de ce texte à la présentation des bases théoriques de la définition 2.

1 - MINIMAX - ROBUSTESSE.

Sur (R, B) soit une famille \mathcal{B} de lois de probabilité. Pour tout $F \in \mathcal{B}$ et tout \mathcal{O} dans R on construit $\mathcal{B}^{\mathcal{O}} = \{F^{\mathcal{O}} \mid F \in \mathcal{B}\}$ où $F^{\mathcal{O}}(x) = F(x - \mathcal{O})$. Soit $T = (T_n)$ une suite d'estimateurs du paramètre de position \mathcal{O} équivariants par translation, définis sur $(R^n, B^{\otimes n})$; la suite T est elle-même élément d'une famille \mathcal{T} . On note encore $\mathcal{B}_{\mathcal{O}} = (\mathcal{B}^{\mathcal{O}})^{\otimes n}$, $\mathcal{B} = \mathcal{B}^{\otimes n}$ où par exemple $\mathcal{B}^{\otimes n} = \{P_1 \times \dots \times P_n \mid P_i \in \mathcal{B}\}$. Parmi les nombreuses mesures des performances de T sur $\mathcal{B}_{\mathcal{O}}$ deux seront étudiées dans les divers chapitres de cette monographie.

A n fixé, une mesure de la plus mauvaise performance liée à l'intervalle de confiance est :

$$(1) \quad S(\mathcal{B}, T_n) = \sup \{ \max [P^{\mathcal{O}}\{T_n(\mathcal{O} - a)\}, P^{\mathcal{O}}\{T_n > \mathcal{O} + a\}] \mid P^{\mathcal{O}} \in \mathcal{B}_{\mathcal{O}} \} .$$

Puisque T_n est équivariant par translation, $S(\mathcal{B}, T_n)$ ne dépend pas de \mathcal{O} . Pour tout $P^{\mathcal{O}}$ de $\mathcal{B}_{\mathcal{O}}$ l'intervalle de confiance bilatéral de longueur fixe $2a$, $[T_n - a, T_n + a]$ aura un niveau inférieur à $2 S(\mathcal{B}, T_n)$ et les deux intervalles unilatéraux $[T_n - a, +\infty[$ et $]-\infty, T_n + a]$ un niveau inférieur à $S(\mathcal{B}, T_n)$.

Définition 1 :

Une suite d'estimateurs $T^{\mathcal{O}}$ sera dite minimax-robuste parmi la famille \mathcal{T} si et seulement si $S(\mathcal{B}, T_n^{\mathcal{O}}) \leq S(\mathcal{B}, T_n)$ pour tout T_n appartenant à \mathcal{T} (la valeur du paramètre a est fixée).

Dans les études asymptotiques (en particulier pour les suites d'estimateurs convergents asymptotiquement normaux) on utilise (lorsque ces expressions ont un sens)

$$\sigma^2(\mathcal{B}, T) = \sup \{ \lim_{n \rightarrow \infty} \inf_{P^{\mathcal{O}}} \text{Var} \sqrt{n} (T_n - \mathcal{O}) \mid P^{\mathcal{O}} \in \mathcal{B}_{\mathcal{O}} \} .$$

Remarquons que σ^2 ne dépend pas de \mathcal{O} puisque T_n est équivariant par translation.

Définition 2 :

Une suite T^0 sera appelée minimax-robuste pour le critère de la variance asymptotique si et seulement si :

$$\sigma^2(\mathcal{B}, T_0) \leq \sigma^2(\mathcal{B}, T) \text{ pour tout } T \in \mathcal{T}$$

Remarques :

1 - Dans les deux cas, asymptotique et non asymptotique, la pire des performances d'une statistique minimax-robuste (sur \mathcal{B} pour la famille \mathcal{T}) est meilleure que la pire performance de toute autre statistique de \mathcal{T} . Les statistiques minimax-robustes sont optimales sur \mathcal{T} en ce sens qu'elles minimisent un critère spécifié. Si le domaine \mathcal{B} est réduit à un point, la statistique minimax-robuste est la statistique optimale de la famille \mathcal{T} .

2 - Conflit robustesse-efficacité (HAMPEL [34]).

Une statistique T^0 minimax-robuste pour optimale qu'elle soit dans la famille \mathcal{T} sur \mathcal{B} , peut avoir des performances mauvaises sous le modèle de base (en général inclus dans \mathcal{B}). L'étude des propriétés de T^0 sous le modèle de base et la confrontation avec la statistique optimale usuelle sous ce modèle, permettent d'éclairer le conflit robustesse-efficacité. Cette étude est menée dans le Chapitre IV.

3 - Le choix de l'ensemble de lois \mathcal{T} c'est-à-dire celui de \mathcal{B} est précisé dans chaque étude. La détermination des estimateurs robustes pour la définition 1 est faite dans le Chapitre VIII et pour la définition 2 dans le Chapitre IV.

2 - π -ROBUSTESSE : DISTANCE DE PROHOROV ET PROPRIÉTÉS.

Dans les paragraphes qui suivent le contre-modèle est constitué de l'ensemble des lois d'une boule pour la distance de PROHOROV, centrée au modèle de base. Les avantages de cette modélisation méritent attention et discussion et sont partiellement développés en fin de Chapitre II où est aussi définie la distance de PROHOROV.

Le calcul effectif de cette distance entre deux lois données présente des difficultés certaines. Son utilisation dans les démonstrations de ce texte repose sur un théorème dû à STRASSEN [90] que nous rappelons sous forme de lemme :

Lemme 1 : Soit dans Ω^2 la bande diagonale de largeur δ :

$$\text{Bande}^\delta(\Omega^2) = \{ (\omega, \omega') \mid d(\omega, \omega') \leq \delta \}.$$

P et Q étant deux lois de probabilités sur (Ω, \mathcal{A}) une condition nécessaire et suffisante pour que $\pi(P, Q) \leq \delta$ est qu'il existe une loi de probabilité S sur $(\Omega^2, \mathcal{A}^{\otimes 2})$ de marginales P et Q et telle que $S(\text{Bande}^\delta(\Omega^2)) \geq 1 - \delta$.

Dans tout ce qui suit on posera $\Omega = \mathbb{R}^n$. La distance d choisie dans cette étude est celle du Sup des valeurs absolues des différences des coordonnées.

Lemme 2 : Soient ω et ω' deux échantillons éléments de \mathbb{R}^n , $\tilde{\omega}$ et $\tilde{\omega}'$ les fonctions de répartition associées :

$$d(\omega, \omega') \leq \delta \Rightarrow \pi(\tilde{\omega}, \tilde{\omega}') \leq \delta.$$

Preuve : Ce résultat se démontre en considérant sur $(\mathbb{R}^2, \mathcal{B}^{\otimes 2})$ la loi uniforme dont le support est composé des n couples (ω_i, ω'_i) construits à partir de $\omega = (\omega_1, \dots, \omega_n)$ et $\omega' = (\omega'_1, \dots, \omega'_n)$.

Soit \mathcal{Q} le groupe des permutations des coordonnées de \mathbb{R}^n ; \mathbb{R}^n/\mathcal{Q} l'espace des orbites de \mathcal{Q} dans \mathbb{R}^n .

$\tilde{\mathbb{R}}^n = (\mathbb{R}^n / \mathcal{Q})$ est mis en bijection avec \mathfrak{X}_n ensemble des lois discrètes sur \mathbb{R} dont

les masses sont proportionnelles à $1/n$.

L'application mesurable h de \mathbb{R}^n dans $\tilde{\mathbb{R}}^n$ définie par :

$$h(\omega) = \tilde{\omega} = \{g(\omega) \mid g \in \mathcal{G}\}$$

transporte la loi de probabilité P sur $(\mathbb{R}^n, \mathcal{A})$ en une loi \tilde{P} définie par

$$\tilde{P}(\tilde{A}) = P(h^{-1}(\tilde{A})) .$$

Lemme 3 : Soient P et Q deux lois sur $(\mathbb{R}^n, \mathfrak{B}^{\otimes n})$, \tilde{P} et \tilde{Q} les lois symétrisées sur $(\tilde{\mathbb{R}}^n, \tilde{\mathfrak{B}}^{\otimes n})$.

On a : $\pi(\tilde{P}, \tilde{Q}) \leq \pi(P, Q)$

Preuve : Soient $B \subset \tilde{\mathbb{R}}^n$ et $\bar{B} \subset \mathbb{R}^n$ où $\bar{B} = \{b \in \mathbb{R}^n \mid \tilde{b} \in B\}$.

$$\text{On a } (\bar{B})^\epsilon = \{a \in \mathbb{R}^n, d(a, \bar{B}) \leq \epsilon\} = \overline{B^\epsilon} .$$

En effet si $a \in (\bar{B})^\epsilon$ alors il existe $b \in \bar{B}$ tel que $d(a, b) \leq \epsilon$ implique $\pi(\tilde{a}, \tilde{b}) \leq \epsilon$.

$$\text{Si } b \in \bar{B} \text{ alors } \tilde{a} \in B^\epsilon \text{ si et seulement si } \tilde{a} \in \overline{B^\epsilon} .$$

$$\begin{aligned} \text{Soit } \pi(P, Q) = \delta . \text{ Alors } \tilde{P}(B) &= P(\bar{B}) \leq Q((\bar{B})^\delta) + \delta \\ &\leq Q(\overline{B^\delta}) + \delta \\ &= \tilde{Q}(B^\delta) + \delta \end{aligned}$$

donc $\pi(\tilde{P}, \tilde{Q}) \leq \delta$.

Proposition 1 : Soient F et G deux lois de probabilité sur $(\mathbb{R}, \mathfrak{B})$

$$\pi(F, G) \leq \delta \Rightarrow \pi(\tilde{F}^n, \tilde{G}^n) \leq \sqrt{\delta} .$$

Preuve : Soit $B = \text{Bande}^\delta(\mathbb{R}^2)$ et X la variable indicatrice du complémentaire de B .

D'après le lemme 1, $\pi(F, G) \leq \delta$ entraîne l'existence d'une loi S sur $(\mathbb{R}^2, \mathfrak{B}^{\otimes 2})$ telle que $S\{X = 1\} \leq \delta$ et admettant F et G pour lois marginales.

Soient n couples aléatoires $(a_i, b_i) \in \mathbb{R}^2$ indépendants équadistribués selon la loi S et soit $X_i = X(a_i, b_i)$.

D'après l'inégalité de MARKOV

$$(3) \quad S^n \left\{ \frac{\sum X_i}{n} < \sqrt{\delta} \right\} \geq 1 - \frac{E X_i}{\sqrt{\delta}} \geq 1 - \frac{\delta}{\sqrt{\delta}} = 1 - \sqrt{\delta}.$$

Or $\left(\frac{\sum X_i}{n} < \sqrt{\delta} \right) \Rightarrow \pi(\tilde{\omega}_1, \tilde{\omega}_2) \leq \sqrt{\delta}$, $\tilde{\omega}_1$ et $\tilde{\omega}_2$ étant les fonctions de répartition marginales du n -échantillon de couples (a_i, b_i) .

$$(3) \text{ implique : } S^n \{ \pi(\tilde{\omega}_1, \tilde{\omega}_2) \leq \sqrt{\delta} \} \geq 1 - \sqrt{\delta} \text{ soit } S^n(\text{Bande}^{\sqrt{\delta}}(\mathfrak{F}_n \times \mathfrak{F}_n)) \geq 1 - \sqrt{\delta}$$

Les marginales de S^n (en fait de \tilde{S}^n) étant \tilde{F}^n et \tilde{G}^n , on a toujours d'après le lemme 1, $\pi(\tilde{F}^n, \tilde{G}^n) \leq \sqrt{\delta}$.

3 - π -ROBUSTESSE : DÉFINITIONS DE LA π -ROBUSTESSE ET REMARQUES MÉTHODOLOGIQUES.

Soient $(T_n(X_1, \dots, X_n))$ une suite de statistiques invariantes par permutation, P_n un modèle de base et Q_n une loi du contre-modèle (P_n et Q_n sont des lois jointes de (X_1, \dots, X_n)). En un sens encore intuitif on considérera que la suite T_n est robuste en P_n s'il est possible de diminuer la distance entre les lois de T_n sous P_n et Q_n en diminuant la distance de PROHOROV entre P_n et Q_n .

Dans ce qui suit on note L_K la loi de T_n lorsque $L(X_1, \dots, X_n) = K$. De plus remarquons que T_n étant invariant par permutation : $L_K = L_{\tilde{K}}$.

Deux définitions de robustesse (π -robustesse et robustesse) ont été posées par HAMPEL ; elles sont présentées dans le tableau suivant :

	Modèle de base P_n	Contre-modèle Q_n	
(4)	F^n sur R^n	Q sur R^n tel que $\pi(\tilde{F}^n, \tilde{Q}) < \delta$	(T_n) est <u>π-robuste en F^n</u> ssi : $(\forall \epsilon > 0 \exists \delta > 0 \text{ tel que } \forall n$ $\pi(\tilde{F}^n, \tilde{Q}) < \delta \Rightarrow \pi(L_{\tilde{F}^n}, L_{\tilde{Q}}) < \epsilon$
	F^n sur R^n	G^n sur R^n tel que $\pi(F, G) < \delta$	(T_n) est <u>robuste en F^n</u> ssi : $\forall \epsilon > 0 \exists \delta > 0 \text{ tel que } \forall n$ $\pi(F, G) < \delta \Rightarrow \pi(L_{\tilde{F}^n}, L_{\tilde{G}^n}) < \epsilon$

Remarques :

1. Le modèle de base est celui de V.A. indépendantes équidistribuées.
2. La forme du contre-modèle est plus générale dans la π -robustesse (partie supérieure du tableau) que dans la robustesse (partie inférieure) : les V.A. X_i peuvent être non équidistribuées ou non indépendantes.
3. Une suite de statistiques π -robuste est robuste.

En effet d'après la proposition 1 :

$$\forall \epsilon > 0 \quad \exists \delta^2 \quad \pi(F, G) \leq \delta^2 \Rightarrow \pi(\tilde{F}^n, \tilde{G}^n) \leq \delta \Rightarrow \pi(L_{\tilde{F}^n}, L_{\tilde{G}^n}) \leq \epsilon \quad .$$

Par contre la robustesse n'implique par la π -robustesse comme le montre un exemple de HAMPEL([35] , P. 1893).

Utilisations possibles de la propriété de π -robustesse.

Soit (T_n) une suite de statistiques π -robuste en F^n .

Si la loi de T_n est connue sous le modèle de base, alors en notant $K(x) = \tilde{F}^n \{T_n < x\}$, d'après (4) et pour tout élément Q du contre-modèle :

$$\pi(\tilde{F}^n, \tilde{Q}) < \delta \Rightarrow K(x - \epsilon) - \epsilon \leq L_{\tilde{Q}} \{T_n \leq x\} \leq K(x + \epsilon) + \epsilon ,$$

ce qui fournit un encadrement utilisable de la fonction de répartition de T_n sous toute loi du contre-modèle. On en déduit aussi un encadrement des quantiles de $L_{\tilde{Q}}$.

4 - DEUX CONDITIONS SUFFISANTES DE π -ROBUSTESSE.

Condition suffisante 1 (HAMPEL [35] condition B).

Soit (T_n) une suite d'estimateurs. Pour que (T_n) soit π -robuste en F^n il suffit que:

Pour tout $\epsilon > 0$ il existe $(\delta(\epsilon, F) = \delta)$ et une suite (E_n) , $E_n \subset \mathcal{F}_n$ tels que :

$(\forall n : F^n(E_n) > 1 - \epsilon)$ et $(\tilde{\omega} \in E_n \text{ et } (\pi(\tilde{\omega}, \tilde{\omega}') < \delta) \Rightarrow |T_n(\tilde{\omega}) - T_n(\tilde{\omega}')| < \epsilon)$.

Remarque 1 : Cette condition porte sur la différence entre les valeurs de la statistique T_n en des points $\tilde{\omega}$ et $\tilde{\omega}'$ proches.

Remarque 2 : La condition suivante entraîne elle aussi la π -robustesse de T_n :

$$\forall \epsilon > 0 \quad \exists \delta > 0 : \pi(\tilde{\omega}, \tilde{\omega}') < \delta \Rightarrow |T_n(\tilde{\omega}) - T_n(\tilde{\omega}')| < \epsilon.$$

Cette condition est cependant plus forte que la condition suffisante 1 qui ne pose l'implication ci-dessus que pour certains $\tilde{\omega}$: ceux des ensembles E_n .

Démonstration de la condition suffisante 1.

Notons $P = F^n$. Il faut montrer que pour tout $\epsilon_1 > 0$ il existe δ_1 tel que pour tout n : $\pi(\tilde{P}, \tilde{Q}) < \delta_1 \Rightarrow \pi(L_{\tilde{P}}, L_{\tilde{Q}}) < \epsilon_1$.

1. D'après le lemme 1 :

$\pi(\tilde{P}, \tilde{Q}) \leq \delta_1 \Leftrightarrow$ il existe une loi S sur $\mathcal{F}_n \times \mathcal{F}_n$ dont les marginales sont \tilde{P} et \tilde{Q} et telle que $S\{(\tilde{\omega}, \tilde{\omega}') \mid \pi(\tilde{\omega}, \tilde{\omega}') \leq \delta_1\} \geq 1 - \delta_1$.

2. Soit $\epsilon = \epsilon_1/2$; définis par les hypothèses de la condition suffisante 1 et associées à ϵ , existent $\delta(\epsilon, F) = \delta$ et $E_n \subset \mathcal{F}_n$. Soit $\delta_1 = \inf(\epsilon, \delta)$.

Soient les événements A et B :

$A = \{(\tilde{\omega}, \tilde{\omega}') \mid \tilde{\omega} \in E_n\}$. D'après les hypothèses de la condition suffisante 1,

$$S(A) = \tilde{P}(E_n) \geq 1 - \epsilon.$$

$B = \{(\tilde{\omega}, \tilde{\omega}') \mid \pi(\tilde{\omega}, \tilde{\omega}') \leq \delta\}$. Comme $\delta_1 \leq \delta$:

$$S(B) \geq S\{(\tilde{\omega}, \tilde{\omega}') \mid \pi(\tilde{\omega}, \tilde{\omega}') \leq \delta_1\} \geq 1 - \delta_1.$$

Puisque $(\omega, \omega') \in A \cap B \Rightarrow |T_n(\tilde{\omega}) - T_n(\tilde{\omega}')| < \epsilon \Rightarrow |T_n(\tilde{\omega}) - T_n(\tilde{\omega}')| < \epsilon_1$,

on a :

$$S\{|T_n(\tilde{\omega}) - T_n(\tilde{\omega}')| < \epsilon_1\} \geq S(A \cap B) \geq S(B) - S(\bigcap A) = S(B) - \tilde{P}(\bigcap A) > 1 - \delta_1 - \epsilon > 1 - \epsilon_1.$$

3. La relation $S\{|T_n(\tilde{\omega}) - T_n(\tilde{\omega}')| \leq \epsilon_1\} \geq 1 - \epsilon_1$ exprime que $\pi(L_{\tilde{P}}, L_{\tilde{Q}}) \leq \epsilon_1$:

Soit en effet $W_n(\tilde{\omega}, \tilde{\omega}') = (T_n(\tilde{\omega}), T_n(\tilde{\omega}'))$ et S' la loi de W_n . S' est une loi sur \mathbb{R}^2 dont les marginales sont $L_{\tilde{P}}$ et $L_{\tilde{Q}}$ et telle que

$$S'(\text{Bande}^{\epsilon_1}(\mathbb{R}^2)) = S(W_n^{-1}(\text{Bande}^{\epsilon_1}(\mathbb{R}^2))) \geq S\{(\tilde{\omega}, \tilde{\omega}') | T_n(\tilde{\omega}) - T_n(\tilde{\omega}')| < \epsilon_1\} > 1 - \epsilon_1.$$

Condition suffisante 2 de π -robustesse.

Définition 2 : Une définition préalable est nécessaire.

Une fonctionnelle $T : \mathfrak{F} \rightarrow \mathbb{R}$ est continue en $P \in \mathfrak{F}$ si et seulement si

$$\forall \epsilon > 0 \quad \exists \delta(\epsilon, P) = \delta \text{ telle que } \pi(P, Q) < \delta \Rightarrow |T(P) - T(Q)| < \epsilon.$$

La continuité de T entraîne une propriété asymptotique intéressante de la statistique T_n associée.

Soit T une fonctionnelle et T_n sa restriction à \mathfrak{F}_n .

Si T est continue en P , la suite de variables aléatoires T_n converge en probabilité vers $T(P)$ lorsque n tend vers l'infini. Pour une démonstration il est possible de se reporter à HAMPEL [35].

Condition suffisante 2 :

Soit T une fonctionnelle continue en $P \in \mathfrak{F}$.

Si pour tout n , T_n (restriction de T à \mathfrak{F}_n) considérée comme application de \mathbb{R}^n dans \mathbb{R} est continue, alors (T_n) est π -robuste en P .

Preuve :

La démonstration du théorème consiste à prouver que l'on peut construire la suite $(E_n)_{n \in \mathbb{N}}$: pour n assez grand on utilise la continuité de la fonctionnelle T , pour les autres valeurs de n on utilise la continuité de T_n comme fonction de points.

1. Soit un $\epsilon > 0$, T étant continue en P il existe $\delta_0 > 0$ tel que

$$(\pi(Q, P) < 2\delta_0 \Rightarrow |T(Q) - T(P)| < \frac{\epsilon}{2}).$$

De plus, la convergence p.s. de $\tilde{\omega}_n$ vers P entraîne qu'il existe $N_0 \in \mathbb{N}$ tel que

$$\forall n > N_0 \quad P^n \{ \pi(\tilde{\omega}_n, P) < \delta_0 \} > 1 - \epsilon.$$

On choisit pour $n > N_0$: $E_n = \{ \tilde{\omega}_n \mid \pi(\tilde{\omega}_n, P) < \delta_0 \}$. Pour cette suite on a :

$$P^n(E_n) > 1 - \epsilon \quad \text{et}$$

$$((\tilde{\omega}_n \in E_n) \text{ et } \pi(\tilde{\omega}_n, \tilde{\omega}') < \delta_0) \Leftrightarrow (\pi(\tilde{\omega}_n, P) < \delta_0 \text{ et } \pi(\tilde{\omega}_n, \tilde{\omega}') < \delta_0)$$

$$\Rightarrow (\pi(\tilde{\omega}_n, P) < \delta_0 \text{ et } \pi(\tilde{\omega}', P) < 2\delta_0)$$

$$\Rightarrow (|T_n(\tilde{\omega}_n) - T_n(\tilde{\omega}')| \leq |T_n(\tilde{\omega}_n) - T(P)| + |T_n(\tilde{\omega}') - T(P)| < \epsilon).$$

2. Etudions les valeurs de $n \leq N_0$.

Puisque T_n est continue : $\forall \omega \in \mathbb{R}^n \quad \exists \delta_n(\omega) > 0$:

$$|\omega - \omega'| < \delta_n(\omega) \Rightarrow |T_n(\omega) - T_n(\omega')| < \epsilon.$$

On va construire δ_n et E'_n ($n \leq N_0$) tels que pour tout n , $P^n(E'_n) > 1 - \epsilon$ et tels que si $\omega \in E'_n$, $|\omega - \omega'| < \delta_n \Rightarrow |T_n(\omega) - T_n(\omega')| < \epsilon$.

Soit $U(\omega)$ la boule ouverte de centre ω et de rayon $\delta_n(\omega)/2$. Soit $(\alpha_i)_{i \in \mathbb{N}}$ une suite décroissante vers zéro de réels strictement positifs. Si

$A_i = \{ \omega \mid \delta_n(\omega) > \alpha_i \}$ et $B_i = \bigcup_{\omega \in A_i} U(\omega)$ alors $\mathbb{R}^n = \bigcup_N A_i = \bigcup_N B_i$. Il existe

j tel que $P^n(\bigcup_1^j B_i) > 1 - \epsilon$. Posons $E'_n = \bigcup_1^j B_i$. Alors

$\forall \omega' \in E'_n : |\omega' - \omega''| < \frac{\alpha_i}{2} = \delta_n \Rightarrow |T_n(\omega') - T_n(\omega'')| < \epsilon$. On peut toujours choisir les α_i de sorte que $\delta_n < 1/N_0$.

Pour tout $n \leq N_0$ posons $E_n = \tilde{E}'_n$.

Soit $\tilde{a} \in E_n$, $\tilde{b} \in \mathfrak{F}_n$ tels que $\pi(\tilde{a}, \tilde{b}) < \delta_n$ alors il existe $a \in \mathbb{R}^n$ et $b \in \mathbb{R}^n$ tels que

$$\pi(\tilde{a}, \tilde{b}) < \delta_n \Rightarrow |a - b| < \delta_n \quad (\text{en effet } \delta_n < 1/N_0)$$

Ceci entraîne que $|T_n(\tilde{a}) - T_n(\tilde{b})| = |T_n(a) - T_n(b)| < \epsilon$.

3. En posant $\delta = \inf \{\delta_0, (\delta_n)_{n \leq N_0}\}$ qui est strictement positif on satisfait à la condition suffisante 1.

Avant de passer à l'étude d'exemples donnons quelques propriétés complémentaires liant continuité et robustesse. Pour les démonstrations on peut se reporter à HAMPEL [35].

1. Si (T_n) est robuste en F et converge en probabilité en tout G d'un voisinage de F vers $U(G)$, alors la fonctionnelle U ainsi définie est continue en F .

2. Soit T une fonctionnelle et T_n sa restriction à \mathfrak{F}_n . Alors les conditions suivantes sont équivalentes :

* T est continue en tout $F \in \mathfrak{F}$.

* T est robuste et la suite T_n converge en probabilité vers $T(F)$ en tout F .

5. POINT DE RUPTURE (HAMPEL [35]) . (Pour des valeurs numériques on peut se reporter au tableau 2 du Chapitre XIV).

Soit (T_n) une suite d'estimateurs. On appelle point de rupture de (T_n) en une loi de probabilité $F \in \mathfrak{F}$ le scalaire δ^* défini par : $\delta^* = \sup \{ \delta \leq 1 \mid \text{il existe un compact } K(\delta) \text{ sous-ensemble propre de l'espace des paramètres tel que :}$

$$\pi(F, G) < \delta \Rightarrow \lim_{n \rightarrow \infty} G^n \{T_n \in K\} \rightarrow 1 \} . \quad \text{Cette définition est}$$

légèrement différente de celle donnée dans HAMPEL [34] . Cette notion fournit en particulier une mesure de l'éloignement maximal que l'on peut admettre pour la loi G tout en maintenant T_n asymptotiquement presque sûrement borné.

6. EXEMPLES.

Nous fournissons ici des résultats concernant la π -robustesse des estimateurs étudiés dans cette monographie. Les démonstrations figurent dans les chapitres correspondants.

Exemples 1 :

	Condition	Continuité de la fonctionnelle	π -robustesse
Moyenne	T (F) unique	nulle part	nulle part
Médiane		en F mais pas dans un voisinage de F	en F mais pas dans un voisinage
Moyenne - α tronquée	T (F) unique	partout	partout
Moyenne Huber-Winzorisée		en F	en F

Ces résultats sont déduits des propriétés générales qui suivent.

Exemple 2 : R-estimateurs (Chapitre XI)

Considérons les R-estimateurs T_n associés aux tests $h_n = \sum_{i=1}^n \varphi\left(\frac{R_i}{n}\right) \text{sg } X_i$. Supposons φ croissante, intégrable et strictement positive sur $]0,1[$. Supposons F dérivable, de dérivée f continue, strictement positive sur \mathbb{R} . Alors T_n est π -robuste en F.

Exemple 3 : L-estimateurs (Chapitre XII)

Soit F une loi sur \mathbb{R} ayant une densité f continue et strictement positive. Soit μ une mesure signée sur $[0, 1]$ concentrée sur $[\alpha, \beta]$ pour $0 < \alpha < \beta < 1$ et telle que $\int |d\mu| < \infty$. Soit $T(F) = \int_0^1 F^{-1}(t) d\mu(t)$. Soit $T_n = T \mid M_n$. Alors (T_n) est π -robuste en F.

Exemple 4 : M-estimateurs (Chapitre X)

Les définitions étant celles du chapitre IV , soit une fonction ψ monotone non décroissante ne gardant pas un signe constant et soit F une loi de support \mathbb{R} . Si $\lambda_F(t) = \int \psi(x - t) F(dt)$ est définie, on pose

$$T(F) = \frac{1}{2} [\text{Sup} \{ t \mid \lambda_F(t) > 0 \} + \text{Inf} \{ t \mid \lambda_F(t) < 0 \}] .$$

Les propositions i) et ii) sont alors équivalentes :

- i) T_n est π -robuste en F
- ii) ψ est bornée et $\lambda_F(t) = 0$ admet une solution unique .

Le théorème s'applique en particulier à la moyenne (nulle part robuste), à la médiane, à la moyenne Huber-Winzorisée.

Claude DENIAU
Georges OPPENHEIM
Claude VIANO
UER de Mathématique
Université Paris V
12, rue Cujas
75005 PARIS

CHAPITRE X

ROBUSTESSE DES M-ESTIMATEURS

par

Claude DENIAU, Georges OPPENHEIM, Claude VIANO

0 - INTRODUCTION.

L'objet de ce chapitre est de démontrer :

I) Une condition nécessaire et suffisante sur ψ (définie au Chapitre IV) pour obtenir des estimateurs π -robustes (Chapitre IX) donc robustes.

II) Des conditions suffisantes sur ψ pour obtenir des estimateurs robustes, en ce sens que si $F \in \mathcal{M}$, alors $\sup_{\mathcal{M}} \text{var} \{ \sqrt{n} [T_n - \xi_0] \} < \infty$.

On démontrera le résultat suivant : lorsqu'il existe, le M-estimateur le plus robuste sur \mathcal{M} est l'estimateur du maximum de vraisemblance pour une loi F_0 d'information de FISHER, $I(F_0)$ minimum sur \mathcal{M} (Maximum likelihood type estimators (P.J.HUBER [49])).

Certains problèmes ne seront pas abordés dans ce texte, nous les citons et donnons quelques références :

* le cas multidimensionnel et le cas où σ est inconnu : P.J. HUBER [44], ANDREWS et al. [2] ; RELLES [78] .

* la régression robuste : P.J. HUBER [50] .

* les tests et intervalles de confiances minimax robustes : P.J. HUBER [45] et [48 bis] , HUBER-CAROL [54] ainsi que les Chapitres VII et VIII du présent volume.

1 - LA π -ROBUSTESSE DES M-ESTIMATEURS.

A toute application ψ de \mathbb{R} dans \mathbb{R} , non décroissante, prenant des valeurs positives et négatives, on peut associer une suite (T_n) de M-estimateurs (voir le chapitre IV).

On se donne une loi de probabilité F dont le support est \mathbb{R} , et on se propose d'étudier la π -robustesse de (T_n) en F .

Théorème 1 :

Si ψ et F vérifient les conditions qui précèdent, les propositions i) et ii) sont équivalentes :

- i) T_n est π robuste en F
- ii) ψ est bornée et $\lambda_F(t) = 0$ a une solution unique.

Notons que l'unicité de la solution de $\lambda_F(t) = 0$ signifie que le segment $[T^*(F), T^{**}(F)]$ se réduit à un point ; on peut, en effet, toujours modifier ψ de façon à annuler λ_F en ce point sans que cela modifie la valeur de T_n .

Démonstration de ii) \Rightarrow i). On utilise la condition suffisante 2 du Chapitre IX :

a) T est faiblement continue en F car, a étant fixé, on ne peut avoir en même temps $|T(F) - T(G)| > a$ et $\pi(F, G) < \eta$ si η est suffisamment petit.

En effet $\pi(F, G) < \eta$ entraîne l'existence sur \mathbb{R}^2 d'une loi Q telle que $Q\{|x - y| < \eta\} \geq 1 - \eta$ et dont les marginales sont F et G . On a alors par définition de $T(F)$ et $T(G)$

$$\forall k > 0 \quad S(k) = \int_{\mathbb{R}^2} [\psi(x - T(F)) - \psi(y - T(G) + k)] dQ(x, y) \leq 0$$

$$\text{Or } S = S_1 + S_2 \quad \text{avec} \quad S_1 = \int_{|x-y| \leq \eta} [\psi(x - T(F)) - \psi(y - T(G) + k)] dQ(x, y)$$

$$S_2 = \int_{|x-y| > \eta} [\psi(x - T(F)) - \psi(y - T(G) + k)] dQ(x, y)$$

Supposons $T(G) \geq T(F) + a$ et prenons $\eta < a$ et $k \leq a - \eta$:

$$|x - y| \leq \eta \Rightarrow \psi(x - T(F)) - \psi(y - T(G) + k) \geq \psi(x - T(F)) - \psi(x + a - T(G)) \geq 0$$

ce qui entraîne

$$S_1 \geq \int_{|x-y| \leq \eta} [\psi(x - T(F)) - \psi(x + a - T(G))] dQ = \int_{\mathbb{R}^2} [\psi(x - T(F)) - \psi(x + a - T(G))] dQ \\ - \int_{|x-y| > \eta} [\psi(x - T(F)) - \psi(x + a - T(G))] dQ$$

La première intégrale du membre de droite est strictement positive si, $\lambda_F(t) = 0$ n'a qu'une solution, et la valeur absolue de la deuxième est inférieure à $\eta \sup |\psi|$.

Il s'ensuit donc que dès que η est suffisamment petit $S(k)$ est strictement positive pour $0 < k \leq a - \eta$ ce qui est impossible.

La même démarche permet de constater l'incompatibilité de $T(G) \leq T(F) - a$ avec

$$S'(h) = \int_{\mathbb{R}^2} [\psi(x - T(F)) - \psi(y - T(G) - h)] dQ(x, y) \geq 0 \quad \forall h > 0.$$

La faible continuité de T en F est donc démontrée.

b) Il reste à montrer que $T_n(x_1, \dots, x_n)$ est continue en tant que fonction de $x = (x_1, \dots, x_n)$ c'est-à-dire que $d(x, y) = \sup_i |x_i - y_i| \rightarrow 0 \Rightarrow |T_n(x) - T_n(y)| \rightarrow 0$. Posant $d(x, y) = d$ on a quels que soient i et θ .

$$(1.1) \quad x_i - \theta - d \leq y_i - \theta \leq x_i - \theta + d$$

. Supposons que $T^*(x) = T^{**}(x) = T_n(x)$. En faisant dans (1.1) successivement $\theta = T_n(x) + h$ puis $\theta = T_n(x) - h$ on obtient si $d < h$

$$\sum \psi(y_i - T_n - h) \leq \sum \psi(x_i - T_n + d - h) < 0$$

$$\text{et} \quad \sum \psi(y_i - T_n + h) \geq \sum \psi(x_i - T_n + d + h) > 0$$

donc $d < h \Rightarrow |T_n(x) - T_n(y)| < h$ ce qui prouve la continuité en x de T_n .

. Supposons maintenant $T^*(x) \neq T^{**}(x)$: $\sum \psi(x_i - \theta)$ s'annule en tout point de l'intervalle $]T^*(x), T^{**}(x)[$. Dès que $d < \frac{T^{**}(x) - T^*(x)}{2}$, pour tout t de l'intervalle

$]T^* + d, T^{**} - d[$ l'application de (1.1) à θ donne

$$0 = \sum \psi(x_i - d - t) \leq \psi(y_i - t) \leq \sum \psi(x_i + d - t) = 0$$

ce qui implique que $T^*(y) \leq T^*(x) + d < T^{**}(x) - d \leq T^{**}(x) - d \leq T^{**}(y)$

Il s'ensuit par symétrie que si $d < \frac{T^{**}(y) - T^*(y)}{2}$ on a
 $T^*(x) \leq T^*(y) + d < T^{**}(y) - d \leq T^{**}(x)$. Soit en rassemblant les résultats

$$d < \frac{T^{**}(x) - T^*(x)}{4} \Rightarrow d < \frac{T^{**}(x) - T^*(x)}{2} \text{ et } d < \frac{T^{**}(x) - T^*(y)}{2} \Rightarrow |T_n(x) - T_n(y)| < d$$

ce qui prouve la continuité de T_n en x .

Démonstration de i \Rightarrow ii.

a) Si ψ n'est pas bornée, T_n n'est pas robuste en F . Nous utilisons pour le prouver le comportement asymptotique de T_n sous la loi F et sous une contaminée convenablement choisie.

Comportement de T_n sous la loi F :

$$\begin{aligned} \text{Quel que soit } \epsilon, F^n \{T_n(X) \leq T^{**}(F) + \epsilon\} &\rightarrow 1 \text{ lorsque } n \rightarrow \infty. \text{ En effet} \\ F^n \left\{ \frac{\sum \psi(X_i - T^{**} - \epsilon)}{n} - \lambda_F(T^{**} + \epsilon) \leq -\lambda_F(T^{**} + \epsilon) \right\} &= F^n \{ \sum \psi(X_i - T^{**} - \epsilon) \leq 0 \} \\ &\leq F^n \{T_n(X) \leq T^{**} + \epsilon\} \end{aligned}$$

et le premier terme tend vers 1 par application de la loi des grands nombres et parce que $-\lambda_F(T^{**} + \epsilon) > 0$ quel que soit $\epsilon > 0$.

Posons $G_0 = (1 - \eta)F + \eta \delta_{x_0}$ et supposons que $x \rightarrow +\infty \Rightarrow \psi(x) \rightarrow +\infty$:

Quels que soient η et K on peut trouver x_0 tel que $G_0^n \{T_n(X) > T^{**} + K\} \rightarrow 1$ lorsque $n \rightarrow \infty$, en effet

$$G_0^n \left\{ \sum \frac{\psi(X_i - T^{**} - K)}{n} - \lambda_{G_0}(T^{**} + K) > -\lambda_{G_0}(T^{**} + K) \right\} < G_0^n \{T_n(X) > T^{**} + K\}$$

où $\lambda_{G_0}(t) = (1 - \eta) \lambda_F(t) + \eta \psi(x_0 - t)$. Il existe alors x_0 tel que $\lambda_{G_0}(T^{**} + K) > 0$

et pour cet x_0 le premier terme de l'inégalité tend vers 1 lorsque $n \rightarrow \infty$.

On termine en remarquant que les lois images $\mathcal{L}_F(T_n)$ et $\mathcal{L}_{G_0}(T_n)$ ont alors une distance de PROHOROV qui tend vers 1 lorsque $n \rightarrow \infty$ si on choisit η suffisamment grand alors que $\pi(F, G_0) = \eta$ peut être rendue aussi petite que possible.

b) Il reste à montrer que si $T^*(F) \neq T^{**}(F)$, T_n n'est pas robuste en F .

Si $G_0 = (1 - \eta) F + \eta \delta_{x_0}$. Quel que soit $\epsilon < T^{**} - T^*$,
 $\lambda_{G_0}(T^{**} - \epsilon) = \eta \psi(x_0 - T^{**} + \epsilon)$ et puisque $\psi(x_0 - T^{**} + \epsilon) > 0$ pour x_0 convenablement choisi, quel que soit η il existe x_0 tel que $\lambda_{G_0}(t)$ ne s'annule qu'au-dessus de $T^{**} - \epsilon$ quel que soit ϵ . Donc, sous G_0 , $\lim T_n = \mu \geq T^{**}$ (en probabilité), μ diffère de $T(F)$ de plus de $\frac{T^{**} - T^*}{2} \neq 0$ et la démonstration se termine comme en a).

2 - MINIMAX - ROBUSTESSE.

1. Soit $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ un échantillon de taille n . Nous définirons le modèle de base par $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), G^{\otimes n})$ où G est une mesure de probabilité et un voisinage du modèle de base par $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (F^{\otimes n})_{F \in \mathcal{M}})$ où \mathcal{M} est un voisinage de G dans \mathcal{F} en un sens qui va être précisé.

A toute application ψ de \mathbb{R} dans \mathbb{R} non décroissante, prenant des valeurs positives et négatives on peut associer une suite (T_n) de M-estimateur (voir définition des M-estimateurs).

2. Efficacité asymptotique d'un M-estimateur.

Cas particulier :

a) Sous le modèle base G , on caractérisera la performance d'une suite (T_n) de M-estimateurs d'un paramètre θ par la variance asymptotique (si elle existe) de $\sqrt{n}(T_n - \theta)$ notée $\sigma_{G, \psi}^2$.

Soient (T_n^1) et (T_n^2) deux suites de M-estimateurs (associées respectivement à ψ_1 et à ψ_2). On dira que (T_n^1) est plus performant que (T_n^2) sous le modèle de base G si $\sigma_{G, \psi_1}^2 \leq \sigma_{G, \psi_2}^2$. La meilleure performance que l'on puisse attendre d'une suite (T_n) de M-estimateurs est $\sigma_{G, \psi}^2 = \frac{1}{I(G)}$ (avec $I(G) < \infty$ ce qui n'est pas réalisé pour tout G , mais on peut supposer, comme HAMPEL [34], que G possède cette propriété).

b) Dans un voisinage du modèle de base : nous chercherons à déterminer ψ_0 (donc une suite de M-estimateurs) dont la pire des performances sur \mathcal{M} (obtenue en F_0) est meilleure que celle de tout autre ψ , i.e. tel que :

$$\sigma_{F_0, \psi_0}^2 = \inf_{\psi} \sup_{\mathcal{M}} \sigma_{F, \psi}^2 \quad (2.1)$$

La suite (T_n) de M-estimateurs associés à ψ_0 , sera appelée suite d'estimateurs robustes sur \mathcal{M} .

Exemple : P.J. HUBER [44] .

Posons $\mathcal{M} = \{F \mid F = (1 - \epsilon) G + \epsilon H\}$ où $0 \leq \epsilon < 1$, et G une mesure de probabilité, G fixée à support convexe, avec une densité absolument continue g telle que $-\text{Log } g$ est convexe sur le support de G .

Dans ce cas là il existe (F_0, ψ_0) vérifiant (2.1) où $\psi_0 = -\frac{f'_0}{f_0}$, avec

$$f_0(t) = \begin{cases} (1 - \epsilon) g(t_0) e^{k(t-t_0)} & t \leq t_0 \\ (1 - \epsilon) g(t) & t_0 \leq t < t_1 \\ (1 - \epsilon) g(t_1) e^{-k(t-t_1)} & t \geq t_1 \end{cases}$$

où $k \in \mathbb{R}_+$ est donné, $[t_0, t_1] = \{t ; |g'(t) / g(t)| \leq k\}$ et

$$(1 - \epsilon)^{-1} = \int_{t_0}^{t_1} g(t) dt + \frac{g(t_0) + g(t_1)}{k}$$

La suite (T_n^0) associée à ψ_0 est une suite d'estimateurs de maximum de vraisemblance sous le modèle F_0 .

Cas général :

$$\text{Posons } K(F, \psi) = \frac{[\int \psi'(x) F(dx)]^2}{\int \psi^2(x) F(dx)} \quad (\text{lorsque le second membre existe}).$$

Lorsque F est symétrique et ψ suffisamment régulière, $K(F, \psi)$ est l'inverse de la variance asymptotique $\sigma_{F, \psi}^2$ sous F , de la suite (T_n) . C'est une mesure de l'efficacité asymptotique. Nous donnons au paragraphe 5 de ce chapitre quelques exemples dans lesquels $\sigma_{F, \psi}^2 = [K(F, \psi)]^{-1}$ ainsi que des exemples de solutions du problème minimax.

Soit Ψ l'ensemble des fonctions ψ de \mathbb{R} dans \mathbb{R} monotones non décroissantes prenant des valeurs positives et négatives et telles que $K(F, \psi)$ soit définie quel que soit $F \in \mathcal{M}$.

L'introduction de cette fonction K est essentiellement motivée par le lemme suivant

Lemme 1 :

La fonction $K(F, \cdot)$ est convexe :

$$K[(1 - \alpha) F_1 + \alpha F_2, \cdot] = (1 - \alpha) K(F_1, \cdot) + \alpha K(F_2, \cdot) \quad \text{où } F_1 \text{ et } F_2 \in \mathfrak{M} \\ \text{et } \alpha \in [0, 1]$$

Définition 1 :

Par analogie avec ce qui précède on dira qu'une suite (T_n^1) est plus robuste qu'une suite (T_n^2) sous le modèle $F \in \mathfrak{M}$ si

$$\inf_{\mathfrak{M}} K(F, \psi_1) \geq \inf_{\mathfrak{M}} K(F, \psi_2)$$

où (T_n^1) et (T_n^2) sont deux suites de M-estimateurs respectivement associées à ψ_1 et ψ_2 éléments de Ψ .

Définition 2 :

Une suite (T_n^0) de M-estimateurs associée à $\psi_0 \in \Psi$ sera minimax robuste sur $\mathfrak{M} \times \Psi$, s'il existe $F_0 \in \mathfrak{M}$ telle que :

$$K(F_0, \psi_0) = \inf_{\mathfrak{M}} \sup_{\psi} K(F, \psi) \quad (2.2)$$

3. Généralisation de l'information de FISHER et propriétés.

Nous commençons par plonger \mathfrak{F} dans l'ensemble \mathcal{O} des mesures positives de masses inférieures ou égales à un, sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$; nous supposons que \mathcal{O} est muni de la topologie vague (On rappelle que (F_n) converge vers F au sens de la topologie vague si quel que soit $\varphi \in \mathcal{X}(\mathbb{R})$, $\lim_{n \rightarrow \infty} \int \varphi(x) F_n(dx) = \int \varphi(x) F(dx)$.)

Lemme 2 :

Soit $G \in \mathcal{O}$ et $\mathfrak{M} = \{F \in \mathcal{O} \mid \sup_t |F(t) - G(t)| \leq \epsilon\}$; l'ensemble \mathfrak{M} est vaguement compact dans \mathcal{O} .

Preuve : Par exemple dans DIEUDONNE [25]. L'ensemble \mathfrak{M} ne serait pas vague-

ment compact dans \mathfrak{F} , c'est pourquoi ϑ a été introduit.

Remarque 1 :

Dans la suite, le voisinage \mathfrak{M} du modèle de base G , pour la topologie vague, sera supposé vaguement compact. Le lemme 1 nous fournit un exemple de tel voisinage.

Nous allons donner une définition de l'information qui prolonge l'information de FISHER à l'ensemble ϑ de toutes les mesures de ϑ (P.J. HUBER [44] et [52]).

Définition 3 :

Soit I une application de ϑ dans \mathbb{R} définie par :

$$(3.1) \text{ Pour tout } F \in \vartheta : I(F) = \sup_{\varphi} \frac{[\int \varphi'(x) F(dx)]^2}{\int \varphi^2(x) F(dx)} \text{ où } \begin{cases} \varphi \in \mathcal{X}^1(\mathbb{R}) \\ \text{et} \\ \int \varphi^2(x) F(dx) > 0 \end{cases}$$

On appelle information de la mesure $F \in \vartheta$, le scalaire $I(F)$ défini en (3.1).

Remarque 2 :

Soit H l'application (linéaire) de $\mathcal{X}^1(\mathbb{R})$ dans \mathbb{R} définie par

$$H : \varphi \rightarrow \int \varphi'(x) F(dx)$$

$$\text{Alors } I(F) = \|H\|_{L^2(F)}^2.$$

Théorème 2 :

L'information $I(F)$ est finie, si et seulement si $F \in \vartheta_1$ où ϑ_1 est l'ensemble des éléments $F \in \vartheta$ possédant une densité absolument continue telle que $\frac{f'}{f} \in L^2(F)$.

La restriction de I à ϑ_1 est l'information de FISHER, définie sur l'ensemble des mesures de masse inférieure ou égale à un.

Preuve :

a) Condition suffisante : On suppose que F possède une densité absolument continue telle que $\frac{f'}{f} \in L^2(F)$, alors quel que soit $\varphi \in \mathcal{X}^1(\mathbb{R})$

$$\int \varphi'(x) F(dx) = - \int \varphi(x) f'(x) dx \quad (3.2)$$

L'inégalité de SCHWARZ permet d'écrire :

$$\left[\int \varphi(x) f(x) \frac{f'(x)}{f(x)} dx \right]^2 \leq \int \varphi^2(x) F(dx) \times \int \left(\frac{f'(x)}{f(x)} \right)^2 F(dx)$$

La relation (3.2) entraîne :

$$\frac{\left[\int \varphi'(x) F(dx) \right]^2}{\int \varphi^2(x) F(dx)} \leq \int \left(\frac{f'(x)}{f(x)} \right)^2 F(dx)$$

$$\text{soit : } I(F) \leq \int \left[\frac{f'(x)}{f(x)} \right]^2 F(dx) \quad (3.3)$$

$$\text{dont : } I(F) < \infty .$$

b) Condition nécessaire : Considérons l'application H définie dans la remarque (3.2).

Alors sous l'hypothèse $I(F) < \infty$, c'est une forme linéaire continue sur $\chi^1(\mathbb{R})$.

On peut prolonger (HAHN - BANACH) H en une forme linéaire continue H sur $L^2(F)$.

Or $\tilde{H} \in [L^2(F)]^*$ si et seulement si il existe $h \in L^2(F)$ telle que :

$$\forall \varphi \in L^2(F) : \tilde{H}(\varphi) = \int \varphi(x) h(x) F(dx)$$

$$\text{Posons : } \forall x \in \mathbb{R} \quad f(x) = - \int_{-\infty}^x h(y) F(dy)$$

Montrons que f définie ci-dessus est la densité (absolument continue) de F et que $h = - \frac{f'}{f}$. Or $h \in L^2(F)$ ce qui achèvera la première partie de la démonstration.

Soit $\varphi \in \chi^1(\mathbb{R})$, alors :

$$\begin{aligned} \int \varphi'(x) f(x) dx &= - \iint_{y < x} \varphi'(x) h(y) F(dy) dx \\ &= \int \varphi(y) h(y) F(dy) \quad (\text{Théorème de FUBINI}) \\ &= \int \varphi'(y) F(dy) \quad (\varphi \in \chi^1(\mathbb{R})) \end{aligned}$$

donc sous l'hypothèse $I(F) < \infty$, la mesure F admet une densité absolument continue f , telle que $\frac{f'}{f} = -h \in L^2(F)$ et l'information de FISHER $\int \left(\frac{f'(x)}{f(x)} \right)^2 F(dx)$ est finie.

Théorème 3 :

Soit \mathcal{M} un sous-ensemble vaguement compact de \mathcal{C} . Il existe $F_0 \in \mathcal{M}$ telle que pour tout $F \in \mathcal{M}$: $I(F_0) \leq I(F)$.

Preuve :

Soit $\varphi \in \mathcal{X}^1(\mathbb{R})$ ($\varphi \neq 0$) définissons une application K_φ de \mathcal{M} dans \mathbb{R} par

$$F \rightarrow K_\varphi(F) = \frac{\left[\int \varphi^1(x) F(dx) \right]^2}{\int \varphi^2(x) F(dx)}$$

Elle est vaguement continue.

L'application I est l'enveloppe supérieure de la famille $(K_\varphi)_{\substack{\varphi \in \mathcal{X}^1(\mathbb{R}) \\ \varphi \neq 0}}$ d'applications vaguement continues, elle est vaguement semi-continue inférieurement et atteint donc sa borne inférieure sur tout vaguement compact. Ce qui achève la démonstration.

Théorème 4 :

Soit \mathcal{M} un sous-ensemble vaguement compact de \mathcal{C} et $\mathcal{M}_1 = \{F \in \mathcal{M} \mid I(F) < \infty\}$ alors \mathcal{M}_1 est vaguement dense dans \mathcal{M} . La preuve de ce théorème est basée sur les deux lemmes classiques suivants :

Lemme 2 :

Soient G et F deux éléments de \mathcal{C} et $I(F) < \infty$

$$f * G(y) = \int f(x - y) G(dx)$$

est la densité d'une loi $L \in \mathcal{C}$ telle que $I(L) < \infty$. (La preuve immédiate de ce lemme se trouve par exemple, dans HAJEK-SIDAK [31], p. 17-18).

Lemme 3 (Régularisation)

Soit (f_n) une suite de fonctions intégrables par rapport à la mesure de LEBESGUE, à supports tous contenus dans un compact K , telles que :

i) la suite $(\int |f_n(x)| dx)$ soit bornée ,

$$\text{ii) } \lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n(x) dx = 1$$

$$\text{iii) Pour tout complémentaire } V \text{ d'un voisinage de l'origine : } \lim_{n \rightarrow \infty} \int_V f_n(x) dx = 0.$$

Alors pour toute mesure G , la fonction $\ell_n = f_n * G$ est la densité par rapport à la mesure de LEBESGUE d'une mesure L_n , telle que la suite (L_n) converge vaguement vers G .

(Pour une preuve voir J. DIEUDONNE [25], p. 263-64).

Preuve du théorème 4 :

Elle est une application immédiate des lemmes ci-dessus, sachant que \mathcal{M} est vaguement compact.

Quelques résultats utiles :

i) L'application I de \mathcal{C} dans \mathbb{R} est convexe ; c'est une conséquence du résultat suivant :

Soit v_1 et v_2 des éléments de \mathbb{R}_+ et $\alpha \in [0, 1]$, alors :

$$\frac{[\alpha u_1 + (1 - \alpha) u_2]^2}{\alpha v_1 + (1 - \alpha) v_2} \leq \frac{u_1^2}{v_1} + (1 - \alpha) \frac{u_2^2}{v_2} \quad (3.4)$$

ii) L'ensemble $\mathcal{M}_1 = \{F \in \mathcal{M} \mid I(F) < \infty\}$ est convexe.

iii) Une mesure sous stochastique F_0 telle que $I(F_0) = \inf_{F \in \mathcal{M}} I(F)$, où \mathcal{M} est vaguement compact, est un élément de \mathcal{M}_1 .

Caractérisation de F_0 telle que $I(F_0) = \inf_{\mathcal{M}} I(F)$

L'application I étant convexe la détermination de F_0 se ramènera à celle d'un minimum local pour I .

Posons $F_\epsilon = (1 - \epsilon) F_0 + \epsilon F_1$ où $\epsilon \in [0, 1]$ et $F_1 \in \mathcal{M}_1$.

Une C.N.S. pour que I soit minimum en $F_0 \in \mathcal{M}_1$ est que

$$\left[\frac{d}{d\epsilon} I(F_\epsilon) \right]_{\epsilon=0} \geq 0 \quad \text{quel que soit } F_1 \in \mathcal{M}_1.$$

(Dans ce qui suit $f'_\epsilon(x)$, $f'_1(x)$, $f'_0(x)$ sont des dérivées par rapport à x).

$$\begin{aligned} \left[\frac{d}{d\epsilon} I(F_\epsilon) \right]_{\epsilon=0} &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [I(F_\epsilon) - I(F_0)] \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int \left[\frac{f'_\epsilon{}^2}{f_\epsilon}(x) - \frac{f'_0{}^2}{f_0}(x) \right] dx \end{aligned}$$

L'inégalité (3.4) permet d'écrire :

$$0 \leq \frac{1}{\epsilon} \left[\frac{f'_\epsilon{}^2}{f_\epsilon}(x) - \frac{f'_0{}^2}{f_0}(x) \right] \leq \frac{f'_1{}^2}{f_1}(x) - \frac{f'_0{}^2}{f_0}(x)$$

Par application du théorème de la convergence dominée

$$\left[\frac{d}{d\epsilon} I(F_\epsilon) \right]_{\epsilon=0} = \lim_{\epsilon \rightarrow 0} \left[\frac{f'_\epsilon{}^2}{f_\epsilon}(x) - \frac{f'_0{}^2}{f_0}(x) \right] dx$$

Quel que soit $F_1 \in \mathcal{M}_1$:

$$\left[\frac{d}{d\epsilon} I(F_\epsilon) \right]_{\epsilon=0} = \int \left[2 \frac{f'_0}{f_0}(x) (f'_1(x) - f'_0(x)) - \left(\frac{f'_0}{f_0} \right)^2 (x) (f_1(x) - f_0(x)) \right] dx \geq 0 \quad (3.5)$$

Enfin si $\frac{f'_0}{f_0}$ est absolument continue l'expression (3.5) s'écrit :

$$\left[\frac{d}{d\epsilon} I(F_\epsilon) \right]_{\epsilon=0} = \int \left(-2 \left(\frac{f'_0}{f_0}(x) \right)' - \frac{f'_0(x)}{f_0(x)} (f'_1(x) - f'_0(x)) \right) dx$$

Quel que soit $F_1 \in \mathcal{M}_1$:

$$\left[\frac{d}{d\epsilon} I(F_\epsilon) \right]_{\epsilon=0} = \int 4 \frac{(f'_0)^{1/2}}{f_0^{1/2}}(x) (f'_1(x) - f'_0(x)) dx = \int \left[I(F_0) + 4 \frac{(f'_0)^{1/2}}{f_0^{1/2}}(x) \right] f'_1(x) dx \leq 0 \quad (3.6)$$

(On montre de la même manière que la fonction $I(F_\epsilon)$ est dérivable quel que soit $\epsilon \in [0, 1]$).

4. Unicité de F_0 .

Théorème 5 :

Si F_0 est telle que $I(F_0) = \inf_{F \in \mathcal{M}_1} I(F)$ et si la densité f_0 de F_0 est strictement positive, alors F_0 est unique.

Preuve :

Supposons qu'il existe F_0 et $F_1 \in \mathcal{M}_1$ telles que

$$I(F_0) = I(F_1) = \inf_{\mathcal{M}_1} I(F)$$

Alors la convexité de $I(F)$ implique que pour tout $\epsilon \in [0, 1] : I(F_\epsilon) = I(F_0)$;
on peut aussi supposer que F_0 et F_1 sont mutuellement absolument continues (s'il n'en était pas ainsi on pourrait prendre $F_{1/4}$ et $F_{3/4}$).

$$\text{On a } \frac{d^2}{d\epsilon^2} I(F_\epsilon) = 0$$

$$\text{Or } \frac{d^2}{d\epsilon^2} \left(\frac{(f'_\epsilon)^2}{f_\epsilon} \right) = \frac{2}{f_\epsilon^3} (f'_1 f_0 - f'_0 f_1)$$

$$\text{donc } f'_0 f_1 - f'_1 f_0 = 0 \quad (F_1, F_0) \text{ presque partout}$$

$$\frac{f'_0}{f_0} = \frac{f'_1}{f_1} \quad (F_1, F_0) \text{ presque partout}$$

$$\text{Log } f_1 = \text{Log } C f_0 \quad C \in \mathbb{R}_+$$

$$f_1 = C f_0$$

Mais l'hypothèse $I(F_0) = I(F_1)$ implique que $C = 1$ ce qui achève la démonstration.

Nous disposons maintenant de tous les éléments pour résoudre notre problème.

Théorème 6 :

a) Soit \mathcal{M} un sous-ensemble vaguement compact de \mathcal{P} , \mathcal{M}_1 l'ensemble des éléments de \mathcal{M} qui ont une information finie et $F_0 \in \mathcal{M}_1$ telle que

$$I(F_0) = \inf_{F \in \mathcal{M}} I(F) < \infty \text{ et } \psi_0 = - \frac{f'_0}{f_0}.$$

Si $\psi_0 = - \frac{f'_0}{f_0} \in \Psi$, alors (F_0, ψ_0) est minimax i.e. :

$$K(F_0, \psi) \leq K(F_0, \psi_0) = I(F_0) \leq K(F, \psi_0)$$

(On rappelle que Ψ est défini en (X.2)).

b) Si (F_0, ψ_0) est minimax et si Ψ contient un multiple non nul de

$$-f'_0/f_0, \text{ alors } I(F_0) = \inf_{F \in \mathcal{M}} I(F) \text{ et } \psi_0 = -f'_0/f_0 \quad (F_0 \text{ p.p.})$$

Preuve :

a) L'inégalité de SCHWARZ permet d'écrire : $K(F, \psi_0) \leq K(F_0, \psi_0)$. Pour démontrer que $K(F_0, \psi_0) \leq K(F, \psi_0)$ il suffit de montrer que $[\frac{d}{d\epsilon} I(F_\epsilon)]_{\epsilon=0} \geq 0$ ce qui vient d'être démontré.

b) Si (F_0, ψ_0) est minimax et si ψ contient un multiple non nul de f'_0/f_0 :

$$I(F_0) = K(F_0, f'_0/f_0) \leq K(F_0, \psi_0) \leq K(F, \psi_0) \leq K(F, f'_0/f_0) \leq I(F) \leq \infty$$

ce qui achève la démonstration.

Le théorème 6 nous donne la solution du problème du minimax dans $\mathcal{M}_1 \subset \mathcal{M}$. Nous allons maintenant énoncer des conditions suffisantes d'existence du minimax sur \mathcal{M} tout entier.

Théorème 7 :

Si \mathcal{M} et \mathcal{M}_1 sont définis comme dans le théorème 5. Soit $F_0 \in \mathcal{M}_1$ telle que $I(F_0) = \inf_{\mathcal{M}} I(F)$; si F_0 est telle que $\Psi_0 = -\frac{f'_0}{f_0} \in \Psi$, absolument continue, telle que ψ'_0 est, soit continue, soit non négative semi-continue supérieurement et telle que $\lim_{|x| \rightarrow \infty} \psi'_0(x) = 0$, alors (F_0, ψ_0) est minimax sur $\mathcal{M} \times \Psi$.

Preuve :

Montrons que pour tout $F \in \mathcal{M}$: $K(F, \psi_0) \geq I(F_0)$. Soit $F \in \mathcal{M}$; il existe une suite F_n d'éléments de \mathcal{M}_1 telle que $\lim_{n \rightarrow \infty} F_n = F$ au sens de la topologie vague (\mathcal{M}_1 est vaguement dense dans \mathcal{M}) .

$$\text{a) Si } \psi'_0 \in \mathcal{C}^0(\mathbb{R}), \text{ alors } \lim_{n \rightarrow \infty} \int \psi'_0(x) F_n(dx) = \int \psi'_0(x) F(dx)$$

(car $K(\mathbb{R})$ est dense dans $\mathcal{C}^0(\mathbb{R})$) .

b) Si ψ'_0 est s.c.s., non négative et telle que $\lim_{|x| \rightarrow \infty} \psi'_0(x) = 0$, alors :

$$\int \psi'_0(x) F(dx) = \inf_h \int h(x) F(dx) = \inf_h \limsup_n \int h(x) F_n(dx) \geq \limsup_n \int \psi'_0(x) F_n(dx)$$

où les fonctions $h \in \mathcal{C}^0(\mathbb{R})$ et sont telles que $h \geq \psi'_0$.

De même

$$\int \psi_0^2(x) F(dx) = \sup_p \int p(x) F(dx) = \sup_p \liminf_n \int p(x) F_n(dx) \leq \liminf_n \int \psi_0^2(x) F_n(dx)$$

où les fonctions $p \in \mathcal{C}^0(\mathbb{R})$ sont telles que : $0 \leq p \leq \psi_0^2$.

De ce qui précède, on conclut par :

$$\begin{aligned} K(F, \psi_0) &\geq \limsup_n K(F_n, \psi_0) \\ &\geq I(F_0) \end{aligned} \quad \text{ce qui achève la démonstration.}$$

5. Exemples .

- Minimax pour $K(F, \psi)$ et pour $\sigma_{F, \psi}^2$ (HUBER [44]) .

Les théorèmes 6 et 7 fournissent une solution du problème du minimax pour $K(F, \psi)$, or les résultats intéressants sont ceux qui portent sur la variance asymptotique

$\sigma_{F, \psi}^2$ de $\sqrt{n}(T_n - \theta)$.

Le théorème 2 du Chapitre IV donne des conditions suffisantes d'existence de la variance asymptotique de $\sqrt{n} T_n(\theta = 0)$; par exemple si :

- i) ψ_0 bornée monotone croissante, continue, non nécessairement continument différentiable.
- ii) $\lambda_F(0) = 0$ (par exemple ψ_0 impaire, F symétrique).
- iii) $\lambda'_{(F)}(0)$ existe et est non nulle.

Dans quels cas $\sigma_{F, \psi}^2 = [K(F, \psi)]^{-1}$?

Enonçons quelques conditions suffisantes (dont les preuves sont immédiates) à ajouter aux conditions i), ii), iii).

1) Quel que soit $F \in \mathfrak{F}_1 = [\lambda'_{(F)}(0)]^2 = [\int \psi'_0(t) F(dt)]^2$, et $\lambda_F(0) = 0$; alors

pour tout $F \in \mathcal{M}_1 : \sigma_{F, \psi}^2 = \frac{1}{K(F, \psi)}$ et le problème du minimax est résolu pour $\sigma_{F, \psi}^2$ sur $\mathcal{M}_1 \times \Psi$.

2) * Quel que soit $F \in \mathcal{M} : \lambda_F(0) = 0$

** ψ'_0 est uniformément continu ;

Alors le problème du minimax pour $\sigma_{F, \psi}^2$ est résolu sur $\mathcal{M} \times \Psi$.

3) Une condition plus faible que 2) et qui permet de résoudre le problème sur $\mathcal{M} \times \Psi$ est la suivante :

* Quel que soit $F \in \mathcal{M} : \lambda_F(0) = 0$

** Quel que soit $F \in \mathcal{M}$ il existe une suite $(F_n)_n$ d'éléments de \mathcal{M}_1 telle que :

* $\lim_{n \rightarrow \infty} F_n = F$ pour la topologie vague

* $\lim \int \psi'_0(t) F_n(dt) \leq -\lambda'_F(0)$.

Exemple 1 (proposé par P.J. HUBER [44]) .

$\mathcal{M}_1 = \mathcal{B}(\phi, \epsilon) = \{F \mid \sup_t |F(t) - \phi(t)| \leq \epsilon, \text{ symétriques, } I(F) < \infty\}$.

P.J. HUBER, détermine dans ce cas F_0 dont la densité (symétrique) est définie par :

$$F_0(t) = \begin{cases} \phi(a) (\cos \frac{1}{2} ca)^{-2} \cos \frac{1}{2} (ct)^2 & 0 \leq t \leq a \\ \phi(t) & a < t < b \\ \phi(b) e^{-b(t-b)} & t \geq b \end{cases} \quad (5.1)$$

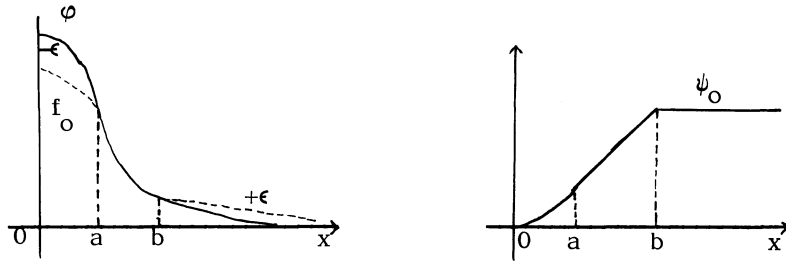
où $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, $\phi(t) = \int_{-\infty}^t \phi(x) dx$ et a, b, c , trois constantes,

fonctions de ϵ , telles que :

$$\text{ctg} \frac{ca}{2} = a \quad (0 \leq ca \leq \pi) \quad (1)$$

$$\int_0^a f_0(t) dt = \int_0^a \phi(t) dt - \epsilon \quad (2) \quad ; \quad \int_b^{\infty} f_0(t) dt = \int_b^{+\infty} \phi(t) dt + \epsilon \quad (3)$$

Principe de la construction de f_0 :



- f_0 symétrique
- $\psi_0 = -\frac{f'_0}{f_0}$ croissante, bornée
- queues alourdies
- F_0 sur la frontière de $\mathcal{B}(\phi, \epsilon)$

La résolution des équations (1), (2), (3) permettant de déterminer a, b, c, n est valable que pour $0 \leq \epsilon < \epsilon_0$ avec ϵ_0 de l'ordre de 0.03.

Notons encore (T_n) la suite des M-estimateurs associées à $\psi_0 = -\frac{f'_0}{f_0}$; alors nous nous trouvons dans des conditions où $\sigma_{F, \psi}^2 = [K(F, \psi)]^{-1}$:

- $\mathcal{L}(\sqrt{n} T_n) \xrightarrow{n \rightarrow \infty} \eta(0, \sigma_{F, \psi_0}^2)$ quel que soit $F \in \mathcal{B}(\phi, \epsilon)$.
- $\sigma_{F, \psi_0}^2 = \frac{1}{K(F, \psi_0)}$ quel que soit $F \in \mathcal{B}(\phi, \epsilon)$ donc (F_0, ψ_0) est minimax pour $\sigma_{F, \psi}^2$.

Exemple 2 : Amélioration du résultat de l'exemple 1. (D. IVILSAKER ; J. SACKS [99])

Il s'agit d'obtenir une solution valable pour $\epsilon > \epsilon_0$ ($\epsilon_0 \simeq 0.03$).

L'idée consiste à utiliser la caractérisation de F_0 , d'information minimum donnée en (3.6).

Plus précisément notons $\mathcal{M}_1 = \{F \mid I(F) < \infty, F(x) + F(-x) = 1\}$.

La mesure $F_0 \in \mathcal{M}_1$ est d'information minimum si et seulement si

$$\forall f \in \mathcal{M}_1 : \int u_0(t) f(t) dt \leq 0$$

$$u_0(t) = I(F_0) + 4 \frac{(f_0^{1/2})''}{f_0^{1/2}}(t) \quad (5.2)$$

Fixons une fonction symétriques $u_0(t)$. Alors F_0 minimise l'information sur $\{F \in \mathcal{M}_1 \mid \int u_0(t) f(t) dt \leq 0\}$. En résolvant l'équation différentielle (5.2) on détermine (F_0, ψ_0) , donc l'estimateur robuste cherché.

Précisons le modèle et posons

$$\mathcal{M}_{A,P} = \{F \in \mathcal{M}_1 \mid \int_{-A}^A f(t) dt \geq p\}$$

il existe $\alpha > 0$ et $\beta < 0$ tels que $u_0(t) = \begin{cases} \alpha & \text{si } |t| > A \\ \beta & \text{si } |t| < A \end{cases}$

$$\text{et } \mathcal{M}_{A,P} = \{F \in \mathcal{M}_1 \mid \int u_0(t) f(t) dt \leq 0\}.$$

On détermine alors f_α solution de (5.2), en posant $\beta(\alpha) = \alpha \tan \alpha$

$$f_\alpha(t) = \begin{cases} \frac{\beta(\alpha)}{1+\beta(\alpha)} \cos^2 \alpha x & \text{si } |x| \leq 1 \\ \frac{\beta(\alpha)}{1+\beta(\alpha)} \cos^2 \alpha e^{2\beta} e^{-2\beta|x|} & \text{si } |x| > 1 \end{cases} \quad (5.3)$$

$$\text{et } \psi_0 = -\frac{f'_\alpha}{f_\alpha}.$$

Considérons $\mathcal{B}(\phi, \epsilon)$ définie dans l'exemple 1 précédent, alors on peut énoncer les résultats suivants :

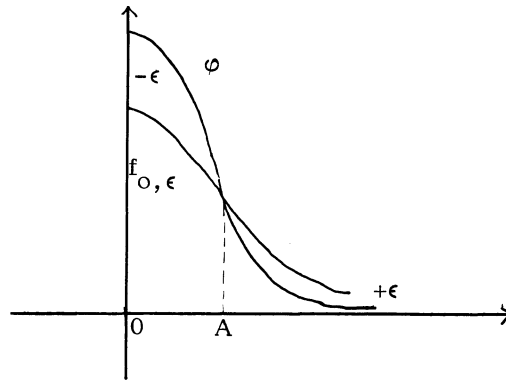
i) Quel que soit $\epsilon \in [0, 1]$ il existe A et p tels que

$$\mathcal{B}(\phi, \epsilon) \subset \mathcal{M}_{A,P}$$

ii) Soit $F_{0,\epsilon}$ mesure d'information minimum dans $\mathcal{M}_{A,P}$; alors $F_{0,\epsilon} \in \mathcal{B}(\phi, \epsilon)$.

iii) Pour $\epsilon > \epsilon_0$ ($\epsilon_0 \simeq 0.03$) $f_{0,\epsilon}$ est définie comme f_α en (5.3).

Remarque : Pour $\epsilon \leq \epsilon_0$, $f_{0, \epsilon}$ est donnée par la formule de HUBER (5.1).



ϕ est la densité de $\mathcal{N}(0, 1)$

Claude DENIAU
 Georges OPPENHEIM
 Claude VIANO
 UER de Mathématiques
 Université Paris V
 12, rue Cujas
 75005 PARIS.

CHAPITRE XI

ROBUSTESSE DES R-ESTIMATEURS

par

Robert AZENCOTT.

Notations : Nous reprenons les notations et définitions du Chapitre V (R-estimateurs).

1 - TOLÉRANCE DES VALEURS EXTRÊMES.

Soit $\hat{\theta}_n$ une suite d'estimateurs du paramètre θ ; pour chaque n , $\hat{\theta}_n$ est une fonction des observations X_1, \dots, X_n . HODGES a proposé dans [38] d'évaluer quantitativement la sensibilité de la suite $\hat{\theta}_n$ à d'éventuelles erreurs grossières, en calculant la proportion maximum de valeurs extrêmes de l'échantillon $X_1 \dots X_n$ qui peuvent subir des perturbations arbitrairement grandes sans affecter trop brutalement $\hat{\theta}_n$.

1.1. Définition :

Pour $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, appelons $y(x) = (y_1(x), \dots, y_n(x))$ le point de \mathbb{R}^n dont les coordonnées se déduisent de celle de x par permutation et tel que $y_1 \leq y_2 \leq \dots \leq y_n$.

Soit U une partie quelconque de \mathbb{R}^n et soit $f : U \rightarrow \mathbb{R}$ une fonction arbitraire.

Nous dirons que f est de tolérance $\tau \geq 0$ s'il existe un entier $k \in [0, \frac{n}{2} - 1[$ vérifiant $\tau = \frac{k}{n}$ et ayant les propriétés suivantes :

- (i) $y_{k+1}(x) \leq f(x) \leq y_{n-k}(x)$ pour tout $x \in U$
- (ii) quels que soient les nombres, $b_{k+2} \leq \dots \leq b_{n-k-1}$, les conditions $y_j(x) = b_j$ pour $k+2 \leq j \leq n-k-1$, $x \in U$, et $y_{n-k}(x) \rightarrow +\infty$ (resp : et $y_{k+1}(x) \rightarrow -\infty$) impliquent $f(x) \rightarrow +\infty$ (resp : $f(x) \rightarrow -\infty$).

Par exemple, pour $U = \mathbb{R}^n$ et $f(x) = \frac{x_1 + \dots + x_n}{n}$, la tolérance de f est égale à 0. Si $g(x)$ est la médiane de $y(x)$, et si $U = \mathbb{R}^n$ la tolérance de g est égale à $\frac{1}{2} - \frac{2}{n}$, pour $n \geq 4$.

1.2. Définitions (d'après HODGES [38]) :

Soit $\hat{\theta}$ un estimateur du paramètre θ , de la forme $\hat{\theta} = f(X_1, \dots, X_n)$ où $X_1 \dots X_n$ sont les observations et f est une fonction mesurable définie sur \mathbb{R}^n . Nous dirons que $\hat{\theta}$ est de tolérance $\tau \geq 0$, si pour chaque θ , il existe une partie borélienne U_θ de \mathbb{R}^n , vérifiant $P_\theta \{(X_1, \dots, X_n) \in U_\theta\} = 1$ telle que $f : U_\theta \rightarrow \mathbb{R}$ soit de tolérance $\tau \geq 0$ (définition 1.1).

Si $\hat{\theta}_n = f_n(X_1, \dots, X_n)$ est une suite d'estimateurs de θ , telle que la tolérance τ_n de $\hat{\theta}_n$ existe pour n assez grand, on appelle tolérance asymptotique τ de la suite $\hat{\theta}_n$, le nombre $\tau = \liminf_{n \rightarrow +\infty} \tau_n$.

La moyenne est donc de tolérance asymptotique 0 et la médiane de tolérance asymptotique $\frac{1}{2}$. Nous allons étendre aux R-estimateurs le calcul de tolérance asymptotique fait pour l'estimateur de HODGES-LEHMANN dans [38]. Les hypothèses et notations sont celles du Chapitre V (R-estimateurs).

1.3. Proposition :

Soit $\hat{\theta}_n$ la suite de R-estimateurs associés aux tests $h_n = \sum_{i=1}^n \varphi\left(\frac{R_i}{n}\right) \text{sign } X_i$.

Supposons φ croissante, continue positive, intégrable sur $]0,1[$. Alors la tolérance asymptotique de $(\hat{\theta}_n)$ est $\geq \alpha$ où $\alpha \in]0, \frac{1}{2}]$ est défini par

$$\int_0^{1-\alpha} \varphi(x) dx = \frac{1}{2} \int_0^1 \varphi(x) dx.$$

Preuve :

Reprenons les notations de V-1-5.2. Soit $k \in [1, n]$. L'ensemble des $M_{j,\ell}$ tels que $n \geq j \geq \ell \geq k$ est de mesure π_n pour $\nu = \nu_{X_1 \dots X_n}$ où

$$\pi_n = \frac{1}{c_n} \left[\sum_{k \leq j \leq n} \sum_{k \leq \ell \leq j} d_1 + j - \ell \right]$$

$$d_i = \varphi(i/n) - \varphi[(i-1)/n] \quad \text{si} \quad i \geq 2 \quad d_1 = \varphi(1/n) \quad c_n = \sum_{i=1}^n \varphi(i/n)$$

ce qui donne
$$\pi_n = \frac{1}{c_n} \left[\sum_{1 \leq z \leq n+1-k} \varphi(z/n) \right]$$

Comme $j \geq k, \ell \geq k$ forcent $M_{j,\ell} \geq M_{kk}$ on voit que $\nu\{[M_{kk}, +\infty[\} \geq \pi_n$. Prenons $k = \beta n, \beta \in]0, \frac{1}{2}]$. On a alors $\lim_{n \rightarrow +\infty} \pi_n = \int_0^{1-\beta} \varphi(x) dx = \pi(\beta)$, en supposant

(ce qui n'est pas une restriction) $\int_0^1 \varphi(x) dx = 1$. Soit α l'unique racine de

$$\pi(\alpha) = \frac{1}{2}.$$

Pour $\beta < \alpha$, on a pour n grand, $\nu([M_{kk}, +\infty[) > \frac{1}{2}$ et donc $M_{kk} = Y_k \leq \hat{\theta}_n$. Un argument analogue montre que $\nu] -\infty, M_{n-k, n-k}] > \frac{1}{2}$ et

$\hat{\theta}_n \leq Y_{n-k}$ pour n grand. On obtient $\tau_n \geq \beta$ pour n grand et donc $\tau \geq \beta$, ce qui donne quand $\beta \nearrow \alpha$, $\tau \geq \alpha$.

Q.E.D.

1.4. Commentaires :

a) Si $\hat{\theta}_n = H.L.$, $\varphi(x) = x$, d'où $\tau \geq 1 - \frac{1}{\sqrt{2}} \simeq 0,29$; en fait dès que $n \geq 4$, $\tau_n \geq 0,25$.

b) Comme α donné par 1.3 est > 0 , si $\varphi > 0$ sur $]01]$, on voit qu'il existe alors $\tau_0 > 0$ tel que $\tau_n \geq \tau_0$ pour tout n .

c) On vérifie aisément que $\tau_n \geq \alpha - \frac{Cte}{n}$ pour tout n si φ est de classe 2 sur $[01]$.

2 - ROBUSTESSE DES R-ESTIMATEURS.

2.1. Nous renvoyons au Chapitre IX pour les définitions de la robustesse et de la π -robustesse d'une suite d'estimateurs. Le fait que les R-estimateurs soient π -robustes est affirmé dans la littérature, mais, à notre connaissance sans démonstration générale, et d'ailleurs sans précisions sur les conditions qui garantissent cette π -robustesse. Nous en fournissons ici une preuve, laborieuse mais complète.

Le modèle de base (suite de variables aléatoires indépendantes X_n , $n \geq 1$, ayant pour loi continue la translatée d'une loi symétrique F par un paramètre $\theta \in \mathbb{R}$) est décrit en détail dans le Chapitre V. Notons que la π -robustesse d'une suite $\hat{\theta}_n$ d'estimateurs de θ traduit une résistance uniforme en n à des perturbations de petite "taille" mais de type très général, qui peuvent fort bien rendre les X_n dépendants, non équidistribués, et de lois non symétriques. Le reste du paragraphe fractionne en plusieurs lemmes la preuve du résultat principal suivant.

2.2. Théorème : (hypothèses et notations du Chapitre V)

Soit φ une fonction croissante, continue, strictement positive sur $]0, 1[$, et intégrable sur $]0, 1[$. Soit $\hat{\theta}_n$, $n \geq 1$ la suite de R-estimateurs associés aux tests
$$h_n = \sum_{i=1}^n \varphi\left(\frac{R_i}{n}\right) \text{sign}(X_i).$$
 Supposons la fonction de répartition F dérivable, de dérivée F' continue strictement positive sur \mathbb{R} . Alors la suite $\hat{\theta}_n$ est π -robuste en F .

2.3. Nous allons utiliser la condition suffisante de π -robustesse due à HAMPEL (cf. Chapitre IX). Soit $M_n(\mathbb{R})$ l'espace des probabilités discrètes sur \mathbb{R} , dont les atomes sont de masses multiples entiers de $\frac{1}{n}$. Soit d la distance de PROKHOROV sur $M_n(\mathbb{R})$. Soit F_n la distribution empirique d'un n -échantillon $X_1 \dots X_n$ sous P_0 (les X_i sont alors indépendants, de même loi F). Alors F_n est une variable aléatoire à valeurs dans $M_n(\mathbb{R})$. Pour tout $\epsilon > 0$, nous allons construire un nombre $\delta > 0$, et une suite \mathcal{S}_n de parties de $M_n(\mathbb{R})$ vérifiant $P_0(F_n \in \mathcal{S}_n) \geq 1 - \epsilon$ pour tout n , et tels

que pour tout n les relations $K \in M_n(\mathbb{R})$, $F_n \in \mathcal{S}_n$ et $d(F_n, K) \leq \delta$ impliquent $|\hat{\theta}_n(F_n) - \hat{\theta}_n(K)| \leq \epsilon$. Ceci prouvera la π -robustesse de la suite $\hat{\theta}_n$.

2.4. Soit $K \in M_n(\mathbb{R})$ et soit $(\bar{X}_1, \dots, \bar{X}_n)$ un point de \mathbb{R}^n tel que K soit la distribution empirique associée à "l'échantillon" $\bar{X}_1, \dots, \bar{X}_n$. D'après le théorème de STRASSEN (cf. [90]) la relation $d(F_n, K) \leq \delta$ signifie que l'on peut passer de l'échantillon X_1, \dots, X_n à $\bar{X}_1, \dots, \bar{X}_n$ comme suit : il existe une partition $J' \cup J''$ de $\{1, \dots, n\}$, avec cardinal $(J'') \leq \delta n$ telle que $|X_i - \bar{X}_i| \leq \delta$ pour tout $i \in J'$. Nous noterons Y_1, \dots, Y_n et $\bar{Y}_1, \dots, \bar{Y}_n$ les échantillons ordonnés associés resp. à X_1, \dots, X_n et $\bar{X}_1, \dots, \bar{X}_n$.

2.5. Lemme :

Pour tout $\epsilon > 0$, et pour tout n fixé, il existe une partie \mathcal{S}_n de $M_n(\mathbb{R})$ vérifiant $P_o(F_n \in \mathcal{S}_n) \geq 1 - \epsilon$, et un nombre $\delta > 0$ tels que les relations $K \in M_n(\mathbb{R})$, $F_n \in \mathcal{S}_n$ et $d(F_n, K) \leq \delta$ impliquent $|\hat{\theta}_n(F_n) - \hat{\theta}_n(K)| \leq \epsilon$.

Preuve :

L'absolue continuité de F entraîne pour $n \geq 2$, $P_o(\inf_{1 \leq i \leq n-1} (Y_{i+1} - Y_i) > 0) = 1$. Etant donnés n et $\epsilon > 0$ il existe donc $a > 0$ tel que

$$P_o(\inf_{1 \leq i \leq n-1} (Y_{i+1} - Y_i) > a) \geq 1 - \epsilon.$$

Soit \mathfrak{F}_n l'ensemble des points de $M_n(\mathbb{R})$ tels que "l'échantillon" ordonné associé $y_1 \dots y_n$ vérifie $\inf_{1 \leq i \leq n-1} (y_{i+1} - y_i) \geq a$. Soit $K \in M_n(\mathbb{R})$ et supposons que $d(F_n, K) \leq \delta$, et $F_n \in \mathfrak{F}_n$. Alors (notations de 2.4), par $\delta < \frac{1}{n}$ on a cardinal $(J'') \leq \delta n < 1$ et donc $J'' = \emptyset$, d'où $|X_i - \bar{X}_i| \leq \delta$ pour tout $i \in [1, n]$. Si $\delta < \frac{a}{3}$ toute relation du type $X_i > X_j$ implique donc $\bar{X}_i > \bar{X}_j$ puisque $|X_i - X_j| \geq a$. On en déduit que $\text{rang}(X_i) = \text{rang}(\bar{X}_i)$ pour $1 \leq i \leq n$ et par suite $|Y_i - \bar{Y}_i| \leq \delta$ pour $1 \leq i \leq n$.

En particulier les nombres $M_{ji} = \frac{1}{2} (Y_j + Y_i)$ et $\bar{M}_{ji} = \frac{1}{2} (\bar{Y}_j + \bar{Y}_i)$ vérifient

$|M_{ji} - \bar{M}_{ji}| \leq \delta$ pour tout $i, j \in [1, n]$. Posons (cf. Chapitre V)

$d_1 = \varphi\left(\frac{i}{n}\right) - \varphi\left(\frac{i-1}{n}\right)$ et soient ν (resp. $\bar{\nu}$) les probabilités sur \mathbb{R} définies par

$$\nu = \frac{1}{c_n} \sum_{j=1}^n \sum_{k=1}^j d_{1+j-k} \delta_{M_{jk}}$$

$$\bar{\nu} = \frac{1}{c_n} \sum_{j=1}^n \sum_{k=1}^j d_{1+j-k} \delta_{\bar{M}_{jk}}$$

où $c_n = \sum_{i=1}^n \varphi\left(\frac{i}{n}\right)$

On sait que (cf. Chapitre V) $\hat{\theta}_n(F_n)$ et $\hat{\theta}_n(K)$ sont respectivement les médianes de ν et $\bar{\nu}$. La relation $|M_{ji} - \bar{M}_{ji}| \leq \delta$ pour tout $i, j \in [1, n]$ entraîne immédiatement $|\hat{\theta}_n(F_n) - \hat{\theta}_n(K)| \leq \delta$ ce qui prouve le lemme à condition de prendre $\delta = \frac{1}{n} \wedge \frac{a}{3} \wedge \epsilon$.

2.6. Notations :

Donnons-nous $n \in \mathbb{N}$, $0 < \alpha < \frac{1}{2}$, $\delta > 0$, $C > 0$. Appelons $\mathcal{S}_n(\alpha, \delta, C)$ l'ensemble des points de $M_n(\mathbb{R})$ dont "l'échantillon" ordonné associé $y_1 \dots y_n$ est formé de n nombres distincts, et vérifie

$$-\delta + \frac{1}{C} \frac{k}{n} \leq y_{i+k} - y_i \leq C \frac{k}{n} + \delta$$

pour $\alpha \leq \frac{i}{n} \leq \frac{i+k}{n} \leq 1 - \alpha$.

2.7. Lemme :

Pour tout $0 < \alpha < \frac{1}{2}$, il existe un nombre $C(\alpha) > 0$ ayant la propriété suivante : quels que soient $\alpha \in]0, \frac{1}{2}[$, $\delta > 0$, $\epsilon > 0$, il existe N tel que $n \geq N$ implique

$$P_o \{F_n \in \mathcal{S}_n(\alpha, \delta, C(\alpha))\} \geq 1 - \epsilon.$$

Preuve :

Soient V_1, \dots, V_n les statistiques d'ordre d'un échantillon de n variables indépendantes U_1, \dots, U_n , de même loi uniforme sur $[0, 1]$, définies sur un espace de

probabilité (Ω, P) . Les variables aléatoires $X_i = F^{-1}(U_i)$ sont indépendantes et de même loi F . L'échantillon ordonné associé est alors donné par $Y_i = F^{-1}(V_i)$. La relative compacité de la suite des lois des variables $\sup_{1 \leq i \leq n} \sqrt{n} |V_i - \frac{i}{n}|$ (cf. [10]) montre que $\epsilon > 0$ étant donné qu'il existe $B > 0$ tel que

$$P \left\{ \sup_{1 \leq i \leq n} |V_i - \frac{i}{n}| \leq \frac{B}{\sqrt{n}} \right\} \geq 1 - \epsilon.$$

Pour tout $\omega \in \Omega$ vérifiant $|V_i(\omega) - \frac{i}{n}| \leq \frac{B}{\sqrt{n}}$, $1 \leq i \leq n$ la formule des accroissements finis et la relation $Y_i = F^{-1}(V_i)$ entraînent

$$\left(-\frac{2B}{\sqrt{n}} + \frac{k}{n}\right) \frac{dF^{-1}}{dx}(z) \leq Y_{i+k}(\omega) - Y_i(\omega) \leq \left(\frac{2B}{\sqrt{n}} + \frac{k}{n}\right) \frac{dF^{-1}}{dx}(z)$$

où z est un nombre dépendant de i, k, ω , mais appartenant à l'intervalle $[V_i, V_{i+k}]$ donc a fortiori à $[\frac{i}{n} - \frac{B}{\sqrt{n}}, \frac{i+k}{n} + \frac{B}{\sqrt{n}}]$, et où $1 \leq i \leq i+k \leq n$.

Soit $\alpha \in]0, \frac{1}{2}[$ fixé ; d'après la continuité et la stricte positivité de $\frac{dF^{-1}}{dx}$ sur $]0, 1[$, il existe $C(\alpha) > 0$ tel que $\frac{1}{C(\alpha)} \leq \frac{dF^{-1}}{dx}(y) \leq C(\alpha)$ pour $\frac{\alpha}{2} \leq y \leq 1 - \frac{\alpha}{2}$. En particulier si $\delta > 0$ est donné, choisissons N tel que $[C(\alpha) + \frac{1}{C(\alpha)}] \frac{2B}{\sqrt{N}} \leq \delta$ et $\frac{B}{\sqrt{N}} \leq \frac{\alpha}{2}$. Pour $n \geq N$ et $\alpha \leq \frac{i}{n} \leq \frac{i+k}{n} \leq 1 - \alpha$

le nombre z ci-dessus appartient à $[\frac{\alpha}{2}, 1 - \frac{\alpha}{2}]$ et donc

$$-\delta + \frac{1}{C(\alpha)} \frac{k}{n} \leq Y_{i+k}(\omega) - Y_i(\omega) \leq C(\alpha) \frac{k}{n} + \delta \text{ pour tout } \omega \text{ vérifiant}$$

$$\sup_{1 \leq i \leq n} |V_i(\omega) - \frac{i}{n}| \leq \frac{B}{\sqrt{n}}.$$

On en conclut que pour $n \geq N$, $\alpha \leq \frac{i}{n} \leq \frac{i+k}{n} \leq 1 - \alpha$ on a

$$P \left\{ -\delta + \frac{1}{C(\alpha)} \frac{k}{n} \leq Y_{i+k} - Y_i \leq \delta + \frac{k}{n} C(\alpha) \right\} \geq 1 - \epsilon \text{ c'est-à-dire}$$

$$P_0 \{F_n \in \mathcal{S}_n(\alpha, \delta, C(\alpha))\} \geq 1 - \epsilon.$$

2.8. Lemme :

Soient X_1, \dots, X_n , et $\bar{X}_1, \dots, \bar{X}_n$ deux points de \mathbb{R}^n . Supposons que les distributions empiriques associées notées F_n et K vérifient $d(F_n, K) \leq \delta$. Supposons que $F_n \in \mathcal{S}_n(\alpha, \delta, C)$ (cf. 2.6). Alors si $\frac{1}{n} \leq C\delta$ et $\delta(17C + 2) \leq \alpha$ les échan-

tillons ordonnés associés Y_1, \dots, Y_n et $\bar{Y}_1, \dots, \bar{Y}_n$ vérifient
 $|Y_i - \bar{Y}_i| \leq 2\delta (6C^2 + C + 1)$ pour $2\alpha \leq \frac{i}{n} \leq 1 - 2\alpha$.

Preuve : (notations de 2.4).

Soit r entier tel que $\frac{r}{n} \in [\beta, 1 - \beta]$ avec $\beta > \alpha$. Soit s en entier tel que $s \geq r$ et $Y_s - Y_r \leq 2\delta$. Si on avait $\frac{s}{n} \geq 1 - \alpha$, il existerait t entier tel que $r \leq t < s$ et $t =$ partie entière de $[n(1 - \alpha)]$. Par suite, on aurait $Y_t - Y_r \leq Y_s - Y_r \leq 2\delta$ avec $\frac{r}{n}, \frac{t}{n} \in [\alpha, 1 - \alpha]$; puisque $F_n \in \mathcal{S}_n(\alpha, \delta, C)$ ceci imposerait (voir 2.6) $-\delta + \frac{1}{C} \frac{t-r}{n} \leq 2\delta$; confrontant avec l'inégalité $\frac{t-r}{n} \geq \beta - \alpha - \frac{1}{n}$, on concluerait que

$$\beta - \alpha \leq \frac{1}{n} + 3C\delta$$

Imposons donc $\frac{1}{n} \leq C\delta$ et choisissons $\beta \geq \alpha + 5C\delta$. On conclut du raisonnement ci-dessus que les relations $\frac{r}{n} \in [\beta, 1 - \beta]$, $s \geq r$ et $Y_s - Y_r \leq 2\delta$ forcent $\frac{s}{n} \in [\alpha, 1 - \alpha]$. Dans les mêmes conditions, on vérifie que $s \leq r$ et $Y_r - Y_s \leq 2\delta$ forcent $\frac{s}{n} \in [\alpha, 1 - \alpha]$.

Alors, d'après (2.6), si $s \in [1, n]$ vérifie $|Y_s - Y_r| \leq 2\delta$, on est certain que

$$\frac{1}{C} \frac{|r-s|}{n} - \delta \leq 2\delta, \text{ d'où } |r-s| \leq 3C\delta n.$$

Ainsi le nombre des Y_s , $s \in [1, n]$ tels que $|Y_r - Y_s| \leq 2\delta$ est majoré par $6C\delta n$.

Supposons toujours $\frac{r}{n} \in [\beta, 1 - \beta]$; supposons de plus que $Y_r = X_i$ avec $i \in J'$. Si $j \in J'$, $X_i - X_j$ et $\bar{X}_i - \bar{X}_j$ ne peuvent être de signes distincts que si $|X_i - X_j| \leq 2\delta$ puisque $|X_i - \bar{X}_i| \leq \delta$ et $|X_j - \bar{X}_j| \leq \delta$. Mais on vient de voir dans une autre terminologie que le nombre des $j \in [1, n]$ tels que $|X_i - X_j| \leq 2\delta$ est majoré par $6C\delta n$. Le cardinal de l'ensemble des $j \in [1, n]$ tels que $X_i - X_j$ et $\bar{X}_i - \bar{X}_j$ soient de signes distincts est donc majoré par $[6C\delta n + \text{card}(J'')]$ donc par $(6C + 1)\delta n$. On en conclut que pour $Y_r = X_i$, $i \in J'$ et $\frac{r}{n} \in [\beta, 1 - \beta]$, le rang \bar{r}

de \bar{X}_i vérifie $|\bar{r} - r| \leq (6C + 1)\delta n$.

Soit $\gamma > \delta$ et soit s un entier tel que $\frac{s}{n} \in [\beta + 2\gamma, 1 - (\beta + 2\gamma)]$. Puisque les X_i sont distincts et $\text{card}(J'') \leq \delta n < \gamma n$, il existe dès que $(\gamma - \delta)n \geq 1$ dans l'intervalle $[s - 2\gamma n, s - \gamma n]$ au moins un entier r tel que $r = \text{rang de } X_i$ avec $i \in J'$. On a alors $\frac{r}{n} \in [\beta, 1 - \beta]$; donc le rang \bar{r} de \bar{X}_i vérifie $|\bar{r} - r| \leq (6C + 1)\delta n$.

Imposons $\gamma \geq (6C + 1)\delta$. Alors on a

$$\bar{r} \leq r + (6C + 1)\delta n \leq s - \gamma n + (6C + 1)\delta n \leq s$$

et donc $\bar{X}_i = \bar{Y}_{\bar{r}} \leq \bar{Y}_s$. Comme $i \in J'$, on sait que $|X_i - \bar{X}_i| \leq \delta$, ce qui entraîne $X_i - \delta \leq \bar{Y}_s$, ou encore puisque $X_i = Y_r$, $Y_r - \delta \leq \bar{Y}_s$. Puisque $\frac{r}{n}$ et $\frac{s}{n}$ appartiennent à $[\alpha, 1 - \alpha]$, et puisque $F_n \in \mathcal{S}_n(\alpha, \delta, C)$, 2.6 et le choix de r prouvent que

$$Y_s - Y_r \leq C \frac{s - r}{n} + \delta \leq 2C\gamma + \delta$$

Par suite on obtient $\bar{Y}_s \geq Y_r - \delta \geq Y_s - 2C\gamma - 2\delta$.

Un argument analogue "à droite de s " permet de prouver que $\bar{Y}_s \leq Y_s + 2C\gamma + 2\delta$ de sorte que finalement $|\bar{Y}_s - Y_s| \leq 2C\gamma + 2\delta$.

En particulier nous pouvons choisir $\beta = \alpha + 5C\delta$, $\gamma = (6C + 1)\delta$, $\frac{1}{n} \leq C\delta$ pour obtenir le résultat : quel que soit s entier tel que

$$\frac{s}{n} \in [\alpha + (17C + 2)\delta, 1 - \alpha - (17C + 2)\delta] \quad \text{on a}$$

$$|\bar{Y}_s - Y_s| \leq 2\delta(6C^2 + C + 1)$$

2.9. Lemme : (notations de 2.4)

Soit $4\tau > 0$ la tolérance asymptotique de la suite d'estimateurs $\hat{\theta}_n$ (cf. 1.2, 1.3, 1.4). Pour tout $\epsilon > 0$ il existe δ, β, γ (nombres > 0) et N tels que si $n \geq N$, les relations $F_n \in \mathcal{S}_n[\tau, \delta, C(\tau)]$ et $|\bar{Y}_i - Y_i| \leq \gamma$ pour $\beta \leq i/n \leq 1 - \beta$ impliquent

P_0 p.s. la relation $|\hat{\theta}_n(F_n) - \hat{\theta}_n(K)| \leq \epsilon$.

Preuve :

Reprenons les notations de la preuve 2.5 pour définir ν , $\bar{\nu}$, d_i , c_n , M_{ij} , \bar{M}_{ij} .
 Supposons que $F_n \in \mathcal{F}_n[\tau, \delta, C]$ et que $|\bar{Y}_i - \bar{Y}_i| \leq \gamma$ pour $\frac{i}{n} \in [\beta, 1 - \beta]$.
 Alors $|M_{ji} - \bar{M}_{ji}| \leq \gamma$ pour $\frac{i}{n}, \frac{j}{n} \in [\beta, 1 - \beta]$.

Si Λ est une partie de $\{(j, i) : 1 \leq i \leq j \leq n\}$ posons

$$E_\Lambda = \{M_{ji}, (j, i) \in \Lambda\} \text{ et } \bar{E}_\Lambda = \{\bar{M}_{ji}, (j, i) \in \Lambda\}.$$

Donnons-nous un entier k , et $\rho \in]0, \tau[$. Posons

$$\Lambda_0 = \{(j, i) : \beta \leq \frac{i}{n} \leq \frac{j}{n} \leq 1 - \beta\},$$

$$\Lambda_- = \{(j, i) : \frac{k}{n} - \rho \leq \frac{i}{n} \leq \frac{j}{n} \leq \frac{k}{n}\}$$

$$\Lambda_+ = \{(j, i) : \frac{k}{n} \leq \frac{i}{n} \leq \frac{j}{n} \leq \frac{k}{n} + \rho\}$$

$$\text{Posons } u_n = \nu [R - E_{\Lambda_0}] = \bar{\nu} [R - \bar{E}_{\Lambda_0}]$$

$$v_n = \nu (E_{\Lambda_+}) = \bar{\nu} (\bar{E}_{\Lambda_+}) \quad (v_n \text{ ne dépend pas de } k).$$

$$\text{Enfin soient } \hat{\theta}_n(F_n) = a_n, \quad \hat{\theta}_n(K) = \bar{a}_n$$

D'après les propriétés de la tolérance asymptotique 4τ de $\hat{\theta}_n$, il existe N tel que $n \geq N$ entraîne P_0 - p.s. la relation $a_n \in [Y_{[3\tau n]}, Y_{[(1-3\tau)n]}]$. Pour chaque n , choisissons donc l'entier (aléatoire) k ci-dessus tel que $\frac{k}{n} \in [2\tau, 1 - 2\tau]$ et

$Y_{k-2} \leq a_n \leq Y_{k-1}$. La relation $F_n \in \mathcal{F}_n(\tau, \delta, C)$ implique alors

$$E_{\Lambda_+} \subset]a_n, a_n + C\rho + C\delta].$$

Comme a_n est médiane de ν , on a $\nu]-\infty, a_n] \geq \frac{1}{2}$ d'où
 $\nu]-\infty, a_n + C\rho + C\delta] \geq \frac{1}{2} + \nu(E_{\Lambda_+}) = \frac{1}{2} + v_n$. Par suite
 $\nu \{]-\infty, a_n + C(\rho + \delta)] \cap E_{\Lambda_0} \} \geq \frac{1}{2} + v_n - u_n$.

Or pour $M_{ji} \in E_{\Lambda_0}$ on a par hypothèse $|M_{ji} - \bar{M}_{ji}| \leq \gamma$, ainsi que

$\nu (M_{ji}) = \bar{\nu} (\bar{M}_{ji})$. La relation $M_{ji} \in]-\infty, a_n + C(\rho + \delta)] \cap E_{\Lambda_0}$ force donc $\bar{M}_{ji} \in]-\infty, a_n + C(\rho + \delta) + \gamma]$ et par suite

$$\bar{\nu}]-\infty, a_n + C(\rho + \delta) + \gamma] \geq \nu \{]-\infty, a_n + C(\rho + \delta)] \cap E_{\Lambda_0} \} \geq \frac{1}{2} + v_n - u_n .$$

Donc si $v_n - u_n > 0$, et si $C(\rho + \delta) + \gamma \leq \epsilon$, on obtient $\bar{\nu}]-\infty, a_n + \epsilon] > \frac{1}{2}$. Alors la médiane \bar{a}_n de $\bar{\nu}$ vérifie $\bar{a}_n \leq a_n + \epsilon$. Le même argument, utilisant Λ_- au lieu de Λ_+ montre que dans les mêmes conditions $\bar{a}_n \geq a_n - \epsilon$, et par suite $|a_n - \bar{a}_n| \leq \epsilon$.

Calculons donc v_n et u_n . Par définition de ν , on a

$$1 - u_n = \frac{1}{c_n} \sum_{1 \leq i \leq n+i-2} [\beta_n] \varphi\left(\frac{i}{n}\right)$$

$$v_n = \frac{1}{c_n} \sum_{1 \leq i \leq 1 + [\rho n]} \varphi\left(\frac{i}{n}\right)$$

En supposant $\int_0^1 \varphi(x) dx = 1$, ce qui n'est pas une restriction évidemment, on voit tout de suite que

$$\lim_{n \rightarrow +\infty} v_n = \int_0^\rho \varphi(x) dx = v(\rho) \quad \text{et que}$$

$$\lim_{n \rightarrow +\infty} u_n = \int_{1-2\beta}^1 \varphi(x) dx = u(\beta) .$$

Etant donné $\epsilon > 0$, fixons donc d'abord $\rho > 0$ tel que $\rho < \tau$ et $\rho C(\tau) \leq \frac{\epsilon}{3}$. Le nombre $v(\rho) > 0$ est alors fixé, choisissons $\beta > 0$ tel que $u(\beta) \leq \frac{1}{2} v(\rho)$. Alors $\lim_{n \rightarrow +\infty} v_n - u_n = v(\rho) - u(\beta) \geq \frac{1}{2} v(\rho) > 0$ et il existe N_1 tel que $n \geq N_1$ implique

$v_n - u_n > 0$. Enfin choisissons δ et γ tels que $C(\tau) \delta \leq \frac{\epsilon}{3}$ et $\gamma \leq \frac{\epsilon}{3}$.

Alors pour $n \geq N \vee N_1$, les relations $F_n \in \mathcal{F}_n(\tau, \delta, C(\tau))$ et $|\bar{Y}_i - Y_i| \leq \gamma$ pour $\frac{i}{n} \in [\beta, 1 - \beta]$, impliquent bien $|\hat{\theta}_n(F_n) - \hat{\theta}_n(K)| = |a_n - \bar{a}_n| \leq \epsilon$.

2.10. Preuve du théorème 2.2 :

Donnons-nous $\epsilon > 0$, φ et F vérifiant les hypothèses de 2.2. Soit (4τ) la tolérance asymptotique des $\hat{\theta}_n$. D'après le lemme 2.9, il existe δ , β , γ , N tels que les relations $F_n \in \mathcal{S}_n[\tau, \delta, C(\tau)]$ et $|\bar{Y}_i - Y_i| \leq \gamma$ pour $\beta \leq \frac{i}{n} \leq 1 - \beta$ impliquent P_o -p.s. la relation $|\hat{\theta}_n(F_n) - \hat{\theta}_n(K)| \leq \epsilon$.

Prenons $\alpha = \frac{\beta}{2}$ et δ_1 tel que

$$2\delta_1 [6C^2(\alpha) + C(\alpha) + 1] \leq \gamma, \quad \delta_1 \leq \frac{\alpha}{17C(\alpha) + 2}$$

D'après le lemme 2.8, les relations $d(F_n, K) \leq \delta_1$, $n \geq 1/C(\alpha)\delta_1 = N'$, et $F_n \in \mathcal{S}_n(\alpha, \delta_1, C(\alpha))$ entraînent $|\bar{Y}_i - Y_i| \leq \gamma$ pour $\beta \leq \frac{i}{n} \leq 1 - \beta$.

D'après le lemme 2.7, il existe N'' tel que $n \geq N''$ entraîne

$$P_o\{F_n \in \mathcal{S}_n(\tau, \delta, C(\tau))\} \geq 1 - \frac{\epsilon}{2} \quad \text{et} \quad P_o\{F_n \in \mathcal{S}_n(\alpha, \delta_1, C(\alpha))\} \geq 1 - \frac{\epsilon}{2}.$$

Posons $\mathcal{F}_n = \mathcal{S}_n(\tau, \delta, C(\tau)) \cap \mathcal{S}_n(\alpha, \delta_1, C(\alpha))$. Alors

$P_o\{F_n \in \mathcal{F}_n\} \geq 1 - \epsilon$ pour $n \geq N''$. Finalement pour $n \geq N \vee N' \vee N'' = N'''$ les relations $F_n \in \mathcal{F}_n$ et $d(F_n, K) \leq \delta_1$ entraînent $|\hat{\theta}_n(F_n) - \hat{\theta}_n(K)| \leq \epsilon$. D'autre part on a $P_o(F_n \in \mathcal{F}_n) \geq 1 - \epsilon$.

Combinons ce résultat avec le lemme 2.5 appliqué à chacun des entiers

$1, 2, \dots, N'''$, pour obtenir la condition suffisante de π -robustesse énoncée en 2.3.

Ainsi la suite $\hat{\theta}_n$ est π -robuste en F .

2.11. Remarque :

Notons que les $\hat{\theta}_n$ auxquels la propriété de π -robustesse s'applique sont les R-estimateurs calculés par la formule de la proposition 5.2, Chapitre V, (médiane de $\nu_{X_1 \dots X_n}$). Nous n'avons pas vérifié l'équivalence de cette formule et de la définition initiale de $\hat{\theta}_n$ ("solution" de $h(X_1 - u, \dots, X_n - u) = 0$) pour des lois sous-

jaçentes non absolument continues (cf. hypothèses de 5.2, Chapitre V) ce qui devrait être fait pour appliquer le théorème 2.2 aux $\hat{\theta}_n$ calculés par l'algorithme de JAECKEL (voir paragraphe 10, Chapitre V).

Robert AZENCOTT
Mathématiques
Université Paris VII
2, Place Jussieu
75221 PARIS CEDEX 05.

CHAPITRE XII

ROBUSTESSE DES L-ESTIMATEURS

par

Catherine HUBER

Soit (X_1, \dots, X_n) un n -échantillon d'une loi sur \mathbb{R} de fonction de répartition F , F_n la fonction de répartition empirique associée, $F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{]-\infty, x]}(X_i)$, $x \in \mathbb{R}$, et ν une mesure signée, définie sur $]0, 1[$ muni de la tribu de ses boréliens, et bornée sur les compacts de $]0, 1[$. La mesure ν définit, sur un sous-ensemble \mathfrak{F}_ν de l'ensemble \mathfrak{F} de toutes les fonctions de répartition sur \mathbb{R} , une fonctionnelle T à valeurs dans \mathbb{R}

$$(1.1) \quad T(G) = \int_0^1 G^{-1}(u) d\nu(u) \quad G \in \mathfrak{F}_\nu.$$

En particulier, pour toute mesure ν du type précédent, \mathfrak{F}_ν contient l'ensemble \mathfrak{F}_n des fonctions de répartition empiriques d'ordre n .

Considérons la suite des statistiques

$$(1.2) \quad T_n = T(F_n) \quad n \in \mathbb{N}.$$

Sous certaines conditions de régularité sur le couple (F, ν) , la suite $(T_n)_{n \in \mathbb{N}}$ est une suite d'estimateurs de $T(F)$ convergente et robuste en F .

Plus précisément, on a le théorème

Théorème.

Soit F une fonction de répartition sur \mathbb{R} et ν une mesure signée bornée à support compact dans $]0,1[$. Si ν ne charge pas les points de discontinuité de F^{-1} , la suite $T_n = \int_0^1 F_n^{-1}(u) d\nu(u)$ $n \in \mathbb{N}$ est une suite d'estimateurs de $T(F)$ convergente et robuste en F .

Remarquons que nous ne nous plaçons pas nécessairement dans le cadre de l'estimation d'un paramètre de translation mais qu'il suffit, pour y revenir, de supposer qu'il existe une fonction de répartition G , symétrique ($G(-x) = 1-G(x)$, $x \in \mathbb{R}$) telle que $F(x) = G(x-\theta)$, $x \in \mathbb{R}$, pour un réel θ , et de considérer des mesures ν sur $]0,1[$ telles que, si on note $G_\theta(x) = G(x-\theta)$, $x \in \mathbb{R}$, $\theta \in \mathbb{R}$, on ait pour tout θ réel $T(G_\theta) = \theta$, ce qui a lieu si ν est symétrique par rapport à $\frac{1}{2}$.

Pour démontrer le théorème, nous aurons besoin du lemme ci-dessous :

Lemme.

Soit ν une mesure signée bornée sur $[0,1]$ et F une fonction de répartition sur \mathbb{R} telles que ν ne charge aucun point de discontinuité de F^{-1} .

Alors, pour tout $\epsilon > 0$; et tout $\epsilon' > 0$, il existe $\eta > 0$ tel que, pour toute fonction de répartition G vérifiant $\pi(F, G) < \eta$, on ait, en notant $|\nu| = \nu^+ + \nu^-$

$$|\nu| \{u \in]0,1[\mid |F^{-1}(u) - G^{-1}(u)| > \epsilon\} < \epsilon'.$$

Démonstration du lemme.

La distance de Prokhorov étant toujours supérieure à la distance de Lévy, si $\pi(F, G) < \eta$, on a

$$F(x-\eta) - \eta \leq G(x) \leq F(x+\eta) + \eta \quad x \in \mathbb{R}.$$

Par symétrie par rapport à la première bissectrice des axes, il en résulte en notant, pour $\eta > 0$ fixé, $F_1(x) = F(x-\eta) - \eta$, $x \in \mathbb{R}$ et $F_2(x) = F(x+\eta) + \eta$, $x \in \mathbb{R}$, que $F_2^{-1}(u) \leq G^{-1}(u) \leq F_1^{-1}(u)$ $u \in]0,1[$.

Or $F_1^{-1}(u) = \inf \{x \mid F_1(x) \geq u\} = \inf \{x \mid F(x-\eta) \geq u+\eta\} = \eta + \inf \{x' \mid F(x') \geq u+\eta\} = \eta + F^{-1}(u+\eta)$. De même, $F_2^{-1}(u) = F^{-1}(u-\eta) - \eta$. Par suite

$$(1.3) \quad F^{-1}(u-\eta) - \eta \leq G^{-1}(u) \leq F^{-1}(u+\eta) + \eta \quad u \in]0, 1[$$

et si $D_\eta(\alpha) = \{u \in]0, 1[\mid |F^{-1}(u) - G^{-1}(u)| > \alpha\}$ et

$$E_\eta(\alpha) = \{u \in]0, 1[\mid |F^{-1}(u+\eta) - F^{-1}(u-\eta) + 2\eta| > \alpha\}, \quad D_\eta(\alpha) \subset E_\eta(\alpha).$$

Quand η décroît vers 0, la suite des ensembles $E_\eta(\alpha)$, pour $\alpha > 0$ fixé, décroît vers l'ensemble des points de discontinuité de F^{-1} où le saut est supérieur à α ; notons $E(\alpha)$ cet ensemble.

Pour tout $\eta > 0$ et tout $\alpha > 0$, $|\nu|(E_\eta(\alpha))$ existe puisque $|\nu|$ est bornée sur $[0, 1]$, et lorsque α restant fixé, η décroît vers 0, $|\nu|(E_\eta(\alpha))$ décroît vers $|\nu|(E(\alpha))$ qui est nulle par hypothèse, ce qui démontre le lemme.

Démonstration du théorème.

D'après le théorème (IX, par. 4), il suffit de démontrer que

a) la fonctionnelle T est continue en F .

b) la fonction qui à tout point (x_1, \dots, x_n) de \mathbf{R}^n fait correspondre $T(F_n)$, où F_n est la fonction de répartition empirique associée à (x_1, \dots, x_n) , est continue sur \mathbf{R}^n muni de la distance ρ

$$\rho((x_1, \dots, x_n), (y_1, \dots, y_n)) = \sup_{i=1, \dots, n} |x_i - y_i|.$$

La propriété a) résulte du lemme précédent et du fait que ν est à support compact :

soit $[\alpha, \beta]$ un segment inclus dans $]0, 1[$ et contenant le support de ν , et notons $|\nu| = \nu^+ + \nu^-$ la somme des parties positive et négative de la mesure signée ν et $\|\nu^+ + \nu^-\|$ la masse totale de $|\nu|$, qui, par hypothèse, est finie. Notons aussi

$$A' = F^{-1}(\alpha^-) \leq A = F^{-1}(\alpha) \leq B = F^{-1}(\beta) \leq B' = F^{-1}(\beta).$$

Soit $\varepsilon > 0$ quelconque.

Il résulte de la démonstration du lemme qu'on peut choisir η assez petit pour que $\pi(F, G) < \eta$ entraîne que $G^{-1}(\alpha) > A' - \varepsilon$ et $G^{-1}(\beta) < B' + \varepsilon$ (inégalité (1.3)) ; soit $\eta = \eta'$, ce nombre.

D'autre part, d'après le lemme, il existe η'' tel que $\pi(F, G) < \eta''$ entraîne que

$$|\nu| \{u \mid \epsilon \leq |F^{-1}(u) - G^{-1}(u)| < \epsilon\} < \epsilon.$$

Choisissons $\eta < \inf(\eta', \eta'')$. Alors, pour tout G tel que $\pi(F, G) < \eta$, on a

$$\begin{aligned} |T(F) - T(G)| &= \left| \int [F^{-1}(u) - G^{-1}(u)] d\nu(u) \right| \leq \int |F^{-1}(u) - G^{-1}(u)| d|\nu|(u) \\ &\leq \int_{\{u \in [\alpha, \beta] \mid |F^{-1}(u) - G^{-1}(u)| \leq \epsilon\}} |F^{-1}(u) - G^{-1}(u)| d|\nu|(u) + \int_{\{u \in [\alpha, \beta] \mid |F^{-1}(u) - G^{-1}(u)| > \epsilon\}} |F^{-1}(u) - G^{-1}(u)| d|\nu|(u) \\ &\leq \epsilon \|\nu^+ + \nu^-\| + \epsilon \cdot 2[\max(|A'|, |B'|) + \epsilon] \end{aligned}$$

qui est un $O(\epsilon)$, ce qui prouve que T est continue en F . En ce qui concerne b), remarquons que si deux points de \mathbb{R}^n , $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$ sont tels que $\rho(x, y) < \epsilon$, les points x' et y' de \mathbb{R}^n , obtenus en ordonnant respectivement les coordonnées de x et celles de y par ordre croissant et notés $x' = (x_{(1)}, \dots, x_{(n)})$, $y' = (y_{(1)}, \dots, y_{(n)})$, ont la même propriété, c'est à dire que $\rho(x', y') < \epsilon$: en effet, si $\inf_{i=1, \dots, n-1} |x_{(i+1)} - x_{(i)}| > 2\epsilon$, la correspondance entre les coordonnées de x' et celles de y' , est la même que celle qu'il y a entre les coordonnées de x et celles de y . Sinon, les interversions ne peuvent avoir lieu que pour des $(x_{(i)}, x_{(i+1)})$ proches à moins de 2ϵ et on a encore dans ce cas $\rho(x', y') < \epsilon$.

Comme T_n s'écrit $T_n = \sum_{i=1}^n X_{(i)} \nu(\frac{i-1}{n}, \frac{i}{n})$ si $X_{(1)}, \dots, X_{(n)}$ sont les statistiques d'ordre du n -échantillon (X_1, \dots, X_n) , il résulte de la remarque ci-dessus et du fait que $|\nu|$ est bornée que, pour tout n , T_n , considérée comme fonction sur \mathbb{R}^n muni de la distance ρ , est continue.

Le théorème est donc démontré. Il admet le corollaire suivant

Corollaire.

Soit ν une mesure signée bornée à support compact dans $]0, 1[$ et absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R} . Alors ν définit sur l'ensemble \mathfrak{F} de toutes les fonctions de répartition sur \mathbb{R} une fonction

$$T(F) = \int_0^1 F^{-1}(t) d\nu(t), \quad F \in \mathfrak{F}, \text{ et si } F_n \text{ est la fonction de répartition empirique}$$

associée à un n -échantillon de la loi de fonction de répartition F , la suite

$T_n = \int_0^1 F_n^{-1}(t) d\nu(t)$, $n \in \mathbb{N}$, est une suite d'estimateurs de $T(F)$ convergente et robuste en F , pour tout F de \mathfrak{F} .

Ce corollaire est utile par exemple dans un cadre non paramétrique, lorsqu'on a besoin d'un estimateur préliminaire, non nécessairement efficace mais ayant de bonnes propriétés de stabilité (autrement dit, de robustesse) et un comportement asymptotique raisonnable, de façon à ce qu'on puisse fonder sur lui, par une modification adaptative appropriée [87], une suite d'estimateurs asymptotiquement efficaces.

Catherine HUBER
Département de Mathématique
Université de Paris-Nord
Avenue J.B. Clément
93430 VILLETANEUSE

CHAPITRE XIII
COMPARAISON DES R-L-M ESTIMATEURS

par

Valentine COSTA, Jean DESHAYES.

INTRODUCTION.

Nous nous proposons de comparer d'après la thèse de L.A. JAEKEL [58] les estimateurs définis dans les chapitres IV, V, VI. Nous disposons toujours d'un n -échantillon X_1, \dots, X_n d'une loi ayant pour fonction de répartition $F_\theta(\cdot) = F(\cdot - \theta)$ où F est supposée symétrique : $F(x) + F(-x) = 1$ pour tout $x \in \mathbb{R}$.

Nous préciserons les hypothèses assurant aux trois types d'estimateurs la normalité asymptotique en nous appuyant sur des théorèmes démontrés dans les chapitres IV, V, et VI, afin de conserver le même critère de comparaison, c'est-à-dire la variance asymptotique.

Nous montrerons, en particulier, qu'il existe une correspondance entre les R, L et M estimateurs, pour F donnée, qui conserve la variance asymptotique et que le résultat de minimax d'HUBER (cf. Chapitre X "Robustesse des M-estimateurs"), s'applique aux trois familles, mais seulement pour des voisinages de contamination symétrique.

I - ÉTUDE COMPARATIVE POUR F DONNÉE.

Les hypothèses choisies sur F sont les suivantes, et dans la suite, on supposera toujours que F les vérifie sauf spécification contraire :

$\mathfrak{H}(F)$: F est symétrique, a une densité f absolument continue, strictement positive sur \mathbb{R} tout entier et $I(F) = \int \left(\frac{f'(x)}{f(x)} \right)^2 f(x) dx < +\infty$.

Alors, $F :]-\infty, +\infty[\rightarrow]0, 1[$ et $F^{-1} :]0, 1[\rightarrow]-\infty, +\infty[$ sont deux bijections inverses, continues. Et $F^{-1}(t) = -F^{-1}(1-t)$ pour tout $t \in]0, 1[$. Remarquons que tous les estimateurs définis en IV, V, VI sont équivariants par translation, aussi étudions-nous leur comportement en $\theta = 0$.

I.1. Définition des trois familles :I.1.1. M-estimateurs :

Soit ψ une fonction réelle telle que

$$\int_{-\infty}^{+\infty} \psi(x - \theta) dF_{\theta}(x) = \int_{-\infty}^{+\infty} \psi(x) dF(x) = 0.$$

Désignons par M_n le M-estimateurs de θ défini par ψ , c'est-à-dire solution de l'équation $\sum_{i=1}^n \psi(X_i - M_n) = 0$.

Nous faisons sur ψ les hypothèses suivantes :

$\mathfrak{H}(M) = \mathfrak{H}(M, F)$: ψ est impaire, croissante, dérivable ; ψ' est à support compact et à variation bornée et $\int_{-\infty}^{+\infty} \psi'(x) dF(x) \neq 0$.

Pour toute fonction ψ vérifiant ces propriétés, nous écrirons $\psi \in \mathfrak{H}(M)$. Ces hypothèses sont suffisantes pour assurer la normalité asymptotique de $\sqrt{n} M_n$ (cf. chapitre IV : "M-estimateurs"). En effet, reprenant les notations du chapitre IV, l'application $\xi \rightarrow \lambda(\xi) = \int_{-\infty}^{+\infty} \psi(x - \xi) dF_{\theta}(x)$ est telle que :

. $\lambda(\theta) = 0$; $\lambda'(\theta) = - \int_{-\infty}^{+\infty} \psi'(x) dF(x)$ existe et $\lambda'(\theta) < 0$. Ceci implique que l'application λ est strictement décroissante au voisinage de 0 ;

. Comme ψ est dérivable et ψ' à support compact, ψ est continue et bornée et donc λ est continue. Ainsi, l'application $\xi \rightarrow \int_{-\infty}^{+\infty} \psi(x - \xi)^2 dF_{\theta}(x)$ est définie et continue en θ .

$$\int_{-\infty}^{+\infty} \psi'(x) dF(x) \neq 0.$$

Nous pouvons conclure que : si $\psi \in \mathcal{H}(M)$, $\sqrt{n} M_n$ converge en loi vers $N(0, \sigma_M^2(F))$ avec

$$\sigma_M^2(F) = \frac{\int_{\mathbf{R}} \psi^2(x) f(x) dx}{\left(\int_{\mathbf{R}} \psi'(x) f(x) dx \right)^2} .$$

I.1.2. L-estimateurs :

Soit $h : [0, 1] \rightarrow \mathbb{R}$ une fonction de poids fixé. Le L-estimateur correspondant est

$$L_n = \frac{1}{n} \sum_{i=1}^n h(i^*) X_{(i)} \quad \text{avec} \quad i^* = \frac{i}{n+1} .$$

Nous faisons sur h les hypothèses suivantes :

$\mathcal{H}_{(L)} : h(t) = h(1-t)$ pour tout $t \in [0, 1]$; $\int_0^1 h(t) dt = 1$; h est à support compact inclus dans $]0, 1[$, et à variation bornée.

Ces hypothèses et les conditions imposées à F assurent la normalité asymptotique de $\sqrt{n} L_n$ (cf. Chapitre VI : L-estimateurs et HUBER [46]) .

En effet, il suffit de vérifier que h et F^{-1} n'ont aucune discontinuité commune sur $]0, 1[$ ce qui est réalisé puisque F^{-1} est continue.

On peut écrire $\sqrt{n} (L_n - \mu)$ converge en loi vers $N(0, \sigma_L^2(F))$ avec

$$\mu = \int_0^1 h(t) F_{\theta}^{-1}(t) dt = \int_0^1 h(t) [F^{-1}(t) + \theta] dt = \theta$$

$\sigma_L^2(F) = \int_0^1 K^2(t) dt - \left(\int_0^1 K(t) dt \right)^2$ où $dK(t) = h(t) dF^{-1}(t)$. Si l'on prend

$$K(t) = \int_{1/2}^t \frac{h(u)}{f \circ F^{-1}(u)} du, \quad \int_0^1 K(t) dt = 0$$

on en conclut que si $h \in \mathcal{H}(L)$, $\sqrt{n} L_n$ converge en loi vers $N(0, \sigma_L^2(F))$ avec

$$\sigma_L^2(F) = \int_0^1 K^2(t) dt \text{ et } K(t) = \int_{1/2}^t \frac{h(u)}{f \circ F^{-1}(u)} du.$$

I.1.3. R-estimateurs :

Nous considérons comme dans le chapitre V, les R-estimateurs définis à l'aide d'une statistique de la forme

$$W_n(X_1, \dots, X_n; r) = \sum_{i=1}^n J\left(\frac{s_i}{2n+1}\right) \quad \text{où}$$

s_i est le rang de $X_i - r$ dans l'ensemble $\{X_1 - r, X_2 - r, \dots, X_n - r, -(X_1 - r), \dots, -(X_n - r)\}$.

Désignons par R_n le R-estimateur défini à l'aide de W_n , c'est-à-dire solution de l'équation $W_n(X_1, \dots, X_n; R_n) = 0$. Nous faisons sur J les hypothèses suivantes :

$\mathcal{H}(R) = \mathcal{H}(R, F) : J : [0, 1] \rightarrow \mathbb{R}$ est croissante ; $J(t) + J(1-t) = 0$

pour tout $t \in [0, 1]$; J est dérivable sur $]0, 1[$ et J' est à support

compact dans $]0, 1[$; $J' \cdot f \circ F^{-1}$ est à variation bornée et

$$\int_0^1 J'(t) \cdot f \circ F^{-1}(t) dt \neq 0.$$

Ces hypothèses assurent d'après le chapitre V, la normalité asymptotique de $\sqrt{n} R_n$.

En effet, en posant $b(i) = J\left(\frac{n+i}{2n+1}\right) = -J\left(\frac{n+1-i}{2n+1}\right)$, on vérifie (cf. chapitre V) que

$$W_n(X_1, \dots, X_n; r) = \sum_{i=1}^n b(\rho_i) \text{ signe}(X_i - r) \quad P_0\text{-p.s.}$$

où ρ_i est le rang de $|X_i - r|$ dans l'ensemble $\{|X_1 - r|, \dots, |X_n - r|\}$

$$\text{Posons } J_n(t) = \begin{cases} J\left(\frac{i}{2n+1}\right) & \frac{i-1}{2n} \leq t < \frac{i}{2n} \quad i = 1, 2, \dots, n \\ 0 & t = \frac{1}{2} \\ J\left(\frac{i}{2n+1}\right) & \frac{i-1}{2n} < t \leq \frac{i}{2n} \quad i = n+1, n+2, \dots, 2n \end{cases}$$

J_n est une fonction en escalier sur $[0, 1]$ telle que $J_n(t) + J_n(1-t) = 0$ pour tout $t \in [0, 1]$, et J_n converge vers J dans $L^2([0, 1])$ lorsque $n \rightarrow +\infty$. Posons $b_n(i) = J_n\left(\frac{n+i}{2n+1}\right) = -J_n\left(\frac{n+1-i}{2n+1}\right)$ alors

$$W_n(X_1, \dots, X_n; r) = \sum_{i=1}^n J_n\left(\frac{s_i}{2n+1}\right) = \sum_{i=1}^n b_n(\rho_i) \text{ Signe}(X_i - r) \quad P_0\text{-p.s.}$$

Les scores $b_n(i)$ ainsi définis pour $i = 1, 2, \dots, n$ sont croissants et la fonction en escalier $\varphi_n : [0, 1] \rightarrow \mathbb{R}$ définie par $\varphi_n(t) = b_n(i)$ pour $t \in \left] \frac{i-n}{n}, \frac{i}{n} \right]$, $i = 1, 2, \dots, n$ et $\varphi_n(0) = 0$ est croissante, et converge vers φ : $\varphi(t) = J\left(\frac{1}{2} + \frac{1}{2}t\right) = -J\left(\frac{1}{2} - \frac{1}{2}t\right)$ pour $t \in [0, 1]$, dans $L^2([0, 1])$. On conclut donc :

Si $J \in \mathcal{H}(R)$, $\sqrt{n}R_n$ converge en loi vers $N(0, \sigma_R^2(F))$ avec :

$$\sigma_R^2(F) = \frac{\int_0^1 J^2(t) dt}{\left[\int_{-\infty}^{+\infty} J' \circ F(x) f(x)^2 dx \right]^2}$$

Remarque : L'hypothèse $\int_0^1 J'(t) \cdot f \circ F^{-1}(t) dt \neq 0$ implique que

$$\int_{-\infty}^{+\infty} J' \circ F(x) f(x)^2 dx \neq 0 \quad \text{puisque} \quad \int_0^1 J'(t) f \circ F^{-1}(t) dt = \int_{-\infty}^{+\infty} J' \circ F(x) f(x)^2 dx$$

Remarque : Il y aurait deux points de vue pour choisir des hypothèses suffisantes assurant la normalité asymptotique : d'une part, imposer de "fortes" restrictions à la loi F des observations et des restrictions "faibles" sur les fonctions ψ, h, J . D'autre part, imposer de "faibles restrictions" à F et des hypothèses fortes sur ψ, h et J . Les deux points de vue ont été développés dans les chapitres IV, V, VI et HUBER [46].

Mais, c'est le second que nous avons choisi car il permet d'établir des correspondances bijectives simples entre les 3 familles et définir des ensembles de fonctions dont les

propriétés dépendent peu de la fonction de répartition F sous-jacente, en particulier $\mathfrak{H}(L)$ n'en dépend pas du tout.

I.1.4. π - robustesse de ces estimateurs:

Les hypothèses choisies sont suffisantes pour que la suite $(M_n)_{n \in \mathbb{N}}$ des π -estimateurs et la suite $(L_n)_{n \in \mathbb{N}}$ des L -estimateurs soient π -robustes en F (chapitres et XI). En effet, la fonction ψ définit une suite $(M_n)_{n \in \mathbb{N}}$ π -robuste en F si et seulement si ψ est bornée et la solution de l'équation $\lambda(\xi) = 0$ est unique. Conditions réalisées si $\psi \in \mathfrak{H}(M)$ car ψ est bornée et $\int_{-\infty}^{+\infty} \psi'(x) dF(x) > 0$

Comme $f(x) > 0$ pour tout $x \in \mathbb{R}$, la suite $(L_n)_{n \in \mathbb{N}}$ définie par une fonction $h \in \mathfrak{H}(L)$ est π -robuste en F .

Si on suppose de plus que $J > 0$ sur l'intervalle $]\frac{1}{2}, 1[$, alors (cf. chapitre XI), la suite $(R_n)_{n \in \mathbb{N}}$ de R -estimateurs définie par $J \in \mathfrak{H}(R)$ est π -robuste en F .

I.2. Comparaison pour $F \in \mathfrak{H}(F)$ donnée.

I.2.1. Définition des applications entre les 3 familles d'estimateurs.

I.2.1.a. Soit $\psi \in \mathfrak{H}(M)$, posons $h(t) = \frac{\psi'(F^{-1}(t))}{\int_{-\infty}^{+\infty} \psi'(x) f(x) dx}$ pour $t \in [0, 1]$.

h est à support compact inclus dans $]0, 1[$ puisque ψ' est à support compact dans \mathbb{R} et $h(t) = h(1-t)$ pour tout $t \in [0, 1]$ puisque ψ' est paire et F symétrique. $\int_0^1 h(t) dt = 1$ et $\psi' \circ F^{-1}$ est à variation bornée puisque ψ' est à variation bornée et F^{-1} continue croissante, donc h est à variation bornée. Ainsi $\psi \xrightarrow{(1)} h$ définit une application de $\mathfrak{H}(M)$ dans $\mathfrak{H}(L)$. Soit $h \in \mathfrak{H}(L)$, posons $\psi(x) = \int_0^x h \circ F(y) dy$ pour $x \in \mathbb{R}$. Comme h est à variation bornée, $h \circ F$ est intégrable au sens de RIE-MANN et ψ est impaire, croissante, dérivable. Sa dérivée $\psi'(x) = h \circ F(x)$ est à support compact. ψ' est bornée et à variation bornée car h est à variation bornée et F croissante. Enfin

$$\int_{-\infty}^{+\infty} \psi'(x) f(x) dx = \int_{-\infty}^{+\infty} h \circ F(x) f(x) dx = \int_0^1 h(t) dt = 1 \quad (2)$$

Donc $\psi \in \mathfrak{H}(M)$ et $h \xrightarrow{(2)} \psi$ définit une application de $\mathfrak{H}(L)$ dans $\mathfrak{H}(M)$. Vérifions que (1) et (2) définissent des "bijections" inverses l'une de l'autre.

$$- \psi \in \mathfrak{H}(M) \xrightarrow{(1)} h_\psi : t \rightarrow h_\psi(t) = \frac{\psi' \circ F^{-1}(t)}{\int_{-\infty}^{+\infty} \psi'(x) f(x) dx}$$

$$h_\psi \xrightarrow{(2)} \psi_{h_\psi} : x \rightarrow \psi_{h_\psi}(x) = \int_0^x h_\psi \circ F(y) dy \quad \text{et} \quad \int_{-\infty}^{+\infty} \psi'_{h_\psi}(x) f(x) dx = 1$$

$$\psi'_{h_\psi}(y) = h_\psi \circ F(y) = \frac{\psi'(y)}{\int_{-\infty}^{+\infty} \psi'(x) f(x) dx}$$

$$\text{Donc} \quad \psi_{h_\psi} = \frac{\psi}{\int_{-\infty}^{+\infty} \psi'(x) f(x) dx} \quad \psi \text{ et } \psi_{h_\psi} \text{ définissent donc le même}$$

M-estimateur.

$$- h \in \mathfrak{H}(L) \xrightarrow{(2)} \psi_h : x \rightarrow \psi_h(x) = \int_0^x h \circ F(y) dy$$

$$\psi_h \xrightarrow{(1)} h_{\psi_h} : t \rightarrow h_{\psi_h}(t) = \frac{\psi'_h \circ F^{-1}(t)}{\int_{-\infty}^{+\infty} \psi'_h(x) f(x) dx}$$

$$\psi'_h(F^{-1}(t)) = h \circ F \circ F^{-1}(t) = h(t) \quad t \in [0, 1]$$

$$\int_{-\infty}^{+\infty} \psi'_h(x) f(x) dx = \int_0^1 h(t) dt = 1 \quad \text{soit } h = h_{\psi_h}$$

I.2.1.b. Soit $\psi \in \mathfrak{H}(M)$, posons $J(t) = \psi(F^{-1}(t))$ pour $t \in [0, 1]$. F^{-1} n'est pas définie aux bornes de l'intervalle $]0, 1[$, mais en posant $F^{-1}(0) = -\infty$, $F^{-1}(1) = +\infty$, la définition de J a un sens car ψ est bornée et $\psi(-\infty) = \inf_{x \in \mathbb{R}} \psi(x) = \lim_{x \rightarrow -\infty} \psi(x)$,

$$\psi(+\infty) = \sup_{x \in \mathbb{R}} \psi(x) = \lim_{x \rightarrow +\infty} \psi(x) = -\psi(-\infty).$$

Pour tout $t \in [0, 1]$, $J(t) + J(1-t) = 0$; J est croissante car ψ et F^{-1} le sont ;

$$J'(t) = \frac{\psi' \circ F^{-1}(t)}{f \circ F^{-1}(t)} \text{ est à support compact inclus dans }]0, 1[.$$

Enfin $J'(t) \cdot f \circ F^{-1}(t) = \psi' \circ F^{-1}(t)$ est à variation bornée et

$$\int_0^1 J'(t) \cdot f \circ F^{-1}(t) dt = \int_0^1 \psi' \circ F^{-1}(t) dt = \int_{-\infty}^{+\infty} \psi'(x) f(x) dx \neq 0.$$

Ainsi $\psi \rightarrow J$ est une application de $\mathcal{H}(M)$ dans $\mathcal{H}(R)$.
(3)

Soit $J \in \mathcal{H}(R)$, posons $\psi(x) = J \circ F(x)$ pour $x \in \mathbb{R}$.

La fonction ψ ainsi définie est impaire et croissante car J et F le sont.

$\psi'(x) = J' \circ F(x) \cdot f(x)$ est à support compact car J' est à support compact dans

$]0, 1[$. ψ' est à variation bornée car $J' \cdot f \circ F^{-1}$ l'est et F est croissante. Enfin

$$\int_{-\infty}^{+\infty} \psi'(x) f(x) dx = \int_0^1 J'(t) \cdot f \circ F^{-1}(t) dt \neq 0.$$

Donc $J \rightarrow \psi$ définit une application de $\mathcal{H}(R)$ dans $\mathcal{H}(M)$.
(4)

Vérifions que (3) et (4) sont deux bijections réciproques.

$$- \psi \in \mathcal{H}(M) \xrightarrow{(3)} J_\psi : t \rightarrow J_\psi(t) = \psi \circ F^{-1}(t)$$

$$J_\psi \xrightarrow{(4)} \psi_{J_\psi} : x \rightarrow \psi_{J_\psi}(x) = J_\psi \circ F(x) = \psi(x)$$

Donc $\psi_{J_\psi} = \psi$

$$- J \in \mathcal{H}(R) \xrightarrow{(4)} \psi_J : x \rightarrow \psi_J(x) = J \circ F(x)$$

$$\psi_J \xrightarrow{(3)} J_{\psi_J} : t \rightarrow J_{\psi_J}(t) = \psi_J \circ F^{-1}(t) = J(t)$$

Soit $J_{\psi_J} = J$.

1.2.1.c. On en déduit les applications suivantes :

$$- h \rightarrow J_h : t \rightarrow J_h(t) = \int_{1/2}^t \frac{h(u)}{f \circ F^{-1}(u)} du$$

$$\mathbb{H}(L) \rightarrow \mathbb{H}(R)$$

$$- J \rightarrow h_J : t \rightarrow h_J(t) = \frac{J'(t) \cdot f \circ F^{-1}(t)}{\int_0^1 J'(t) \cdot f \circ F^{-1}(t) dt}$$

$$\mathbb{H}(R) \rightarrow \mathbb{H}(L)$$

Dans la suite, nous ne spécifierons plus les indices et nous noterons simplement ψ , h et J les fonctions qui se correspondent par ces applications.

1.2.2. Propriétés de ces applications :

Théorème : Les applications conservent la variance asymptotique ; c'est-à-dire, pour des estimateurs images l'un de l'autre par les applications définies ci-dessus :

$$\sigma_M^2(F) = \sigma_R^2(F) = \sigma_L^2(F)$$

Démonstration : (cf. JAECKEL [58])

$$\begin{aligned} \sigma_M^2(F) &= \frac{\int_{\mathbb{R}} \psi^2(x) f(x) dx}{\left(\int_{\mathbb{R}} \psi'(x) f(x) dx \right)^2} \\ \cdot h(t) &= \frac{\psi'(F^{-1}(t))}{\int_{\mathbb{R}} \psi'(x) f(x) dx} \Rightarrow K(t) = \int_{1/2}^t h(u) dF^{-1}(u) \\ &= \frac{1}{\int_{\mathbb{R}} \psi'(x) f(x) dx} \int_{1/2}^t \psi'(F^{-1}(u)) dF^{-1}(u) \\ &= \frac{1}{\int_{\mathbb{R}} \psi'(x) f(x) dx} [\psi(F^{-1}(u))]_{1/2}^t = \frac{\psi(F^{-1}(t))}{\int_{\mathbb{R}} \psi'(x) f(x) dx} \end{aligned}$$

$$\text{Or } \sigma_L^2(F) = \int_0^1 K^2(t) dt = \frac{\int_0^1 [\psi(F^{-1}(t))]^2 dt}{\left(\int_R \psi'(x) f(x) dx\right)^2} = \frac{\int_R \psi(x)^2 f(x) dx}{\left(\int_R \psi'(x) f(x) dx\right)^2} = \sigma_M^2(F)$$

$$\cdot J(t) = \psi(F^{-1}(t)) \Rightarrow \sigma_R^2(F) = \frac{\int_0^1 J(t)^2 dt}{\left(\int J'(F(x)) f^2(x) dx\right)^2}$$

$$= \frac{\int_0^1 [\psi(F^{-1}(t))]^2 dt}{\left(\int \psi'(x) f(x) dx\right)^2} = \sigma_M^2(F) .$$

Conséquence : Pour F donnée, les trois familles sont équivalentes du point de vue asymptotique ; en particulier, si l'une des familles contient un estimateur asymptotiquement efficace pour F , alors les estimateurs correspondant des autres familles le sont aussi.

Cas particulier : Considérons le M-estimateur d'HUBER, défini pour $k > 0$, par

$$\psi(x) = \begin{cases} -k & x < -k \\ x & -k \leq |x| \leq k \\ k & x > k \end{cases}$$

Il définit une moyenne k -Winsorizée (voir P.J. BICKEL [9] et Chapitre IV, M-estimateurs). Le L-estimateur correspondant est défini par :

$$h(t) = \frac{\psi(F^{-1}(t))}{\int \psi'(x) f(x) dx} = \begin{cases} \frac{1}{1 - 2F(-k)} & \text{si } F(-k) < t < F(k) \\ 0 & \text{si } t \leq F(-k) \text{ et } t \geq F(k). \end{cases}$$

C'est une moyenne α -tronquée avec $\alpha = F(-k)$.

L'estimateur d'HUBER qui apparaît, a priori, comme une moyenne Winsorizée est en fait très proche d'une moyenne tronquée. Nous allons en effet montrer que, non seulement les M-estimateurs et L-estimateurs qui se correspondent ont même loi asymptotique, mais de plus, ils sont "très proches en probabilité asymptotiquement".

I.2.3. Proximité des L et M-estimateurs correspondants, pour $F \in \mathcal{H}(F)$:

Lemme : S'il existe des nombres f_0 , ϵ_0 et α_0 , $f_0 > 0$ et $0 < \epsilon_0 < \alpha_0 < \frac{1}{2}$ tels que $f(x) \geq f_0$ pour tout x vérifiant : $\alpha_0 - \epsilon_0 \leq F(x) \leq 1 - \alpha_0 + \epsilon_0$, alors $X_{(i)} - F^{-1}(i^*)$ est $O\left(\frac{1}{\sqrt{n}}\right)$ en probabilité $(O_p\left(\frac{1}{\sqrt{n}}\right))$ uniformément en $i \in I(\alpha_0)$.

$$I(\alpha) = \{ [\alpha n] + 1, [\alpha n] + 2, \dots, n - [\alpha n] \}$$

Remarque : Si f est continue, les hypothèses du lemme sont satisfaites.

Démonstration du lemme : (cf. JAEKEL [58])

D'après un théorème de KOLMOGOROV, la statistique $D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$ est telle que

$$\lim_{n \rightarrow +\infty} P(D_n < \frac{z}{\sqrt{n}}) = 1 - 2 \sum_{k=1}^{+\infty} (-1)^{k+1} e^{-2k^2 z^2}.$$

F_n désigne la fonction de répartition empirique de l'échantillon X_1, \dots, X_n , soit

$$F_n(x) = \begin{cases} \frac{i}{n} & X_{(i)} < x < X_{(i+1)} \\ \frac{i}{n+1} & x = X_{(i)} \end{cases} \quad i = 0, 1, 2, \dots, n$$

en posant $X_{(0)} = -\infty$ $X_{(n+1)} = +\infty$.

Donc, pour tout réel $\delta > 0$, il existe D et un entier $N(\delta)$ tel que :

$$P(\sqrt{n} D_n < D) \geq 1 - \delta, \quad \text{pour tout } n \geq N(\delta).$$

Soit $A_n = \{ \sqrt{n} D_n < D \}$. Sur A_n , nous avons :

$$|F_n(x) - F(x)| \leq \frac{D}{\sqrt{n}} \Rightarrow |i^* - F(X_{(i)})| \leq \frac{D}{\sqrt{n}}$$

pour tout $i = 1, 2, \dots, n$.

Or $\lim_{n \rightarrow +\infty} ([\alpha_0 n] + 1)^* = \alpha_0$ et $\lim_{n \rightarrow +\infty} (n - [\alpha_0 n])^* = 1 - \alpha_0$

Donc on peut trouver un entier N_1 , tel que, pour tout $n \geq N_1$, on ait :

$$\alpha_o - \frac{\epsilon_o}{2} < i^* < 1 - \alpha_o + \frac{\epsilon_o}{2}, \quad \forall i \in I(\alpha_o),$$

et un entier N_2 , tel que pour tout $n \geq N_2$, on ait $\frac{D}{\sqrt{n}} < \frac{\epsilon_o}{2}$.

Alors $\alpha_o - \epsilon_o < F(X_{(i)}) < 1 - \alpha_o + \epsilon_o$, $\forall i \in I(\alpha_o)$, $\forall n \geq \text{Sup}\{N_1, N_2\}$.

On a (théorème des accroissement finis) :

$$F^{-1}(i^*) - F^{-1}(F(X_{(i)})) = [i^* - F(X_{(i)})] \cdot \frac{1}{f(F^{-1}(\eta))}$$

avec $\eta \in [\alpha_o - \epsilon_o, 1 - \alpha_o + \epsilon_o]$

Donc $|F^{-1}(i^*) - X_{(i)}| \leq \frac{D}{\sqrt{n}} \times \frac{1}{f_o} \quad \forall i \in I(\alpha_o)$

Pour tout $n \geq \text{Sup}\{N_1, N_2, N(\delta)\} = N$, cette inégalité est vraie pour une probabilité supérieure ou égale à $1 - \delta$.

Soit $\forall \delta > 0$, $\exists D$, $\exists N$ tels que $\forall n \geq N$:

$$P\{|X_{(i)} - F^{-1}(i^*)| \leq \frac{D}{\sqrt{n} \cdot f_o}, \quad \forall i \in I(\alpha_o)\} \geq 1 - \delta. \quad \text{C.Q.F.D.}$$

Théorème :

. S'il existe des nombres $f_o, \epsilon_o, \alpha_o$, $f_o > 0$ et $0 < \epsilon_o < \alpha_o < \frac{1}{2}$, tels que $f(x) \geq f_o$ pour tout x vérifiant $\alpha_o - \epsilon_o \leq F(x) \leq 1 - \alpha_o + \epsilon_o$, si f est bornée.

. Si $\psi \in \mathcal{H}(M)$, si de plus le support de ψ' est inclus dans l'intervalle $[F^{-1}(\alpha_o), F^{-1}(1 - \alpha_o)]$, ψ' est continue sur son support et ψ'' existe et est bornée sauf en un nombre fini de points, alors $\sqrt{n}(M_n - L_n)$ converge vers 0 en probabilité.

On démontrera en fait que : $M_n - L_n$ est $O\left(\frac{1}{n}\right)$ en probabilité.

Démonstration : On notera M et L au lieu de M_n et L_n au cours de la démonstration.

En appliquant la formule de TAYLOR à la fonction ψ , nous obtenons :

$$\psi(X_{(i)} - M) = \psi(F^{-1}(i^*)) + [X_{(i)} - M - F^{-1}(i^*)] \psi'(F^{-1}(i^*)) + r_i$$

$$\sum_{i=1}^n \psi(X_{(i)} - M) = \sum_{i=1}^n \psi(X_i - M) = 0 \text{ par définition de } M.$$

$$\sum_{i=1}^n \psi(F^{-1}(i^*)) = 0 \quad \text{car} \quad \psi(F^{-1}(1 - i^*)) = \psi(-F^{-1}(i^*)) = -\psi(F^{-1}(i^*))$$

En sommant de $i = 1$ à $i = n$, la première égalité, on obtient :

$$0 = 0 + \sum_{i=1}^n (X_{(i)} - F^{-1}(i^*) - M) \psi'(F^{-1}(i^*)) + \sum_{i=1}^n r_i$$

D'où, en divisant par : $\int_{-\infty}^{+\infty} \psi'(x) f(x) dx = c$

$$0 = \sum_{i=1}^n (X_{(i)} - F^{-1}(i^*) - M) h(i^*) + \frac{1}{c} \sum_{i=1}^n r_i$$

$$\text{Or} \quad \sum_{i=1}^n h(i^*) X_{(i)} = nL \quad \text{et} \quad \sum_{i=1}^n F^{-1}(i^*) h(i^*) = 0$$

à cause des propriétés de symétrie de F^{-1} et h .

Enfin, $\sum_{i=1}^n h(i^*) = n + O(1)$. En effet, puisque le support de h est inclus dans l'intervalle $[\alpha_0, 1 - \alpha_0]$, que h est dérivable (ψ'' existe) sauf en un nombre fini de points et que h est bornée, il existe un nombre positif k tel que

$$\left| \int_0^1 h(t) dt - \frac{1}{n} \sum_{i=1}^n h(i^*) \right| \leq \frac{k}{n}.$$

En reportant, on obtient $0 = nL - nM - O(1)M + \frac{1}{c} \sum_{i=1}^n r_i$; comme $\sqrt{n}M$ converge en loi vers $N(0, \sigma_M^2(F))$, M est bornée en probabilité. Si $\sum_{i=1}^n r_i$ est borné en probabilité, on obtiendra $|n(L - M)| \leq B$ constante, en probabilité, ce qui achèvera la démonstration.

Il reste donc à prouver que $\sum_{i=1}^n r_i$ est borné en probabilité.

$$(1) \quad r_i = [X_{(i)} - F^{-1}(i^*) - M] [\psi' \{F^{-1}(i^*) + t(X_{(i)} - M - F^{-1}(i^*))\} - \psi'(F^{-1}(i^*))]$$

Si $i^* \leq \alpha_0$; et $X_{(i)} - M \leq F^{-1}(\alpha_0)$, alors $r_i = 0$ car ψ est constante pour $x \leq F^{-1}(\alpha_0)$.

Si $i^* \leq \alpha_0$ et $X_{(i)} - M > F^{-1}(\alpha_0)$, on peut supposer :

$i^* \leq \alpha_0$ et $F^{-1}(\alpha_0) < X_{(i)} - M < F^{-1}(1 - \alpha_0)$ car, pour tout i tel que

$i^* \leq \alpha_0$, $X_{(i)} - M \leq X_{(j)} - M$ où j est le plus petit entier tel que $j^* \geq \alpha_0$.

Et on a $|X_{(j)} - F^{-1}(j^*)| < \frac{D}{\sqrt{n} f_0}$ en probabilité.

D'autre part $|M| < \frac{\text{constante}_0}{\sqrt{n}}$ en probabilité.

Pour $i^* \leq \alpha_0$ et $F^{-1}(\alpha_0) \leq X_{(i)} - M \leq F^{-1}(1 - \alpha_0)$, on a :

$$\begin{aligned} \psi(X_{(i)} - M) &= \psi(F^{-1}(\alpha_0)) + [X_{(i)} - M - F^{-1}(\alpha_0)] \psi'(F^{-1}(\alpha_0)) \\ &\quad + \frac{1}{2} [X_{(i)} - M - F^{-1}(\alpha_0)]^2 \psi''(\eta) \end{aligned}$$

$$\text{Et} \quad \psi(X_{(i)} - M) = \psi(F^{-1}(i^*)) + O + r_i$$

$$\text{Donc} \quad r_i = \frac{1}{2} [X_{(i)} - M - F^{-1}(\alpha_0)]^2 \psi''(\eta)$$

$$\text{Et} \quad |r_i| \leq \frac{1}{2} (X_{(i)} - M - F^{-1}(\alpha_0))^2 \text{Sup} |\psi''| \leq \frac{1}{2} (X_{(j)} - M - F^{-1}(\alpha_0))^2 \text{Sup} |\psi''|$$

$$\forall i^* \leq \alpha_0$$

$$\text{Or} \quad X_{(j)} - M - F^{-1}(\alpha_0) = X_{(j)} - F^{-1}(j^*) + F^{-1}(j^*) - F^{-1}(\alpha_0) - M$$

$$X_{(j)} - F^{-1}(j^*) \text{ est } O_p\left(\frac{1}{\sqrt{n}}\right).$$

$$|F^{-1}(j^*) - F^{-1}(\alpha_0)| \leq |j^* - \alpha_0| \frac{1}{f_0} \leq \frac{1}{n \cdot f_0}$$

$$M \text{ est } O_p\left(\frac{1}{\sqrt{n}}\right).$$

Finalement r_i est $O_p\left(\frac{1}{n}\right)$ uniformément en i tel que $i^* \leq \alpha_0$. De même r_i est $O_p\left(\frac{1}{n}\right)$ uniformément en i tel que $i^* \geq 1 - \alpha_0$.

Pour $\alpha_0 < i^* < 1 - \alpha_0$, les $X_{(i)} - F^{-1}(i^*)$ sont $O_p\left(\frac{1}{\sqrt{n}}\right)$ uniformément en i et M est $O_p\left(\frac{1}{\sqrt{n}}\right)$. Donc $[X_{(i)} - M - F^{-1}(i^*)]^2$ est $O_p\left(\frac{1}{n}\right)$ uniformément en i : tel que $i^* \in]\alpha_0, 1 - \alpha_0[$.

Or $r_i = \frac{1}{2} [X_{(i)} - M - F^{-1}(i^*)]^2 \cdot \psi'' \{ F^{-1}(i^*) + t(X_{(i)} - M - F^{-1}(i^*)) \}$

Donc les r_i sont $O_p(\frac{1}{n})$ uniformément en i tel que $i^* \in]\alpha_0, 1 - \alpha_0[$.

Ainsi, si ψ'' existe en tout point et est bornée, les r_i sont $O_p(\frac{1}{n})$ uniformément en $i = 1, 2, \dots, n$. Donc $\sum_{i=1}^n r_i$ est $O_p(1)$, c'est-à-dire borné en probabilité.

Reste à étudier le cas où ψ'' existe sauf en un nombre fini de points. Il suffit de faire la démonstration pour un seul "mauvais" point a . Comme ψ'' n'existe pas au point a , un développement de TAYLOR à l'ordre 2 n'est plus valable pour tous les i tels que $a \in [F^{-1}(i^*), X_{(i)} - M]$ où $a \in [X_{(i)} - M, F^{-1}(i^*)]$ selon le cas. En utilisant la relation (1) et en effectuant une démonstration analogue à celle que nous avons faite ci-dessus, nous saurons seulement que les r_i sont $O_p(\frac{1}{\sqrt{n}})$ uniformément en $i = 1, 2, \dots, n$.

Déterminons donc le nombre $N(a)$ de points i tels que $a \in [X_{(i)} - M, F^{-1}(i^*)]$

$$N(a) = \sum_{i=1}^n 1_{(X_{(i)} - M \leq a \leq F^{-1}(i^*))} + \sum_{i=1}^n 1_{(F^{-1}(i^*) \leq a \leq X_{(i)} - M)}$$

$$N(a) = \sum_{i > (n+1)F(a)} 1_{[X_{(i)} - M \leq a]} + \sum_{i \leq (n+1)F(a)} 1_{[X_{(i)} - M \geq a]}$$

$$N(a) = \sum_{i > (n+1)F(a)} 1_{(X_{(i)} \leq a + M)} + \sum_{i \leq (n+1)F(a)} 1_{(X_{(i)} \geq a + M)}$$

$$N(a) = \sum_{i > (n+1)F(a)} 1_{[\frac{i}{n+1} \leq F_n(a + M)]} + \sum_{i \leq (n+1)F(a)} 1_{[\frac{i}{n+1} \geq F_n(a + M)]}$$

$$\text{car } F_n(X_{(i)}) = \frac{i}{n+1}$$

$$N(a) = \sum_{i > (n+1)F(a)} 1_{[i \leq (n+1)F_n(a+M)]} + \sum_{i \leq (n+1)F(a)} 1_{[i \geq (n+1)F_n(a+M)]}$$

$$N(a) = | [F_n(a + M) - F(a)] | (n + 1)$$

$$\text{Or } \sqrt{n} | F_n(a + M) - F(a) | \leq \sqrt{n} | F_n(a + M) - F(a + M) | + \sqrt{n} | F(a + M) - F(a) |$$

$\sqrt{n} |F_n(a+M) - F(a+M)|$ est $O_p(1)$ d'après le théorème de KOLMOGOROV cité au début. Et $\sqrt{n} |F(a+M) - F(a)| \leq \sqrt{n} M \sup_{n \in \mathbb{R}} f(x)$ est aussi $O_p(1)$. On peut remarquer ici qu'il n'est pas nécessaire de supposer que f est bornée sur \mathbb{R} , mais seulement au voisinage des points où ψ'' a une discontinuité.

Donc $N(a)$ est $O_p(\sqrt{n})$. La contribution dans $\sum_{i=1}^n r_i$ de ces $N(a)$ restes est donc $N(a) \cdot O_p\left(\frac{1}{\sqrt{n}}\right) = O_p(1)$. Donc $\sum_{i=1}^n r_i$ est encore $O_p(1)$.

Remarque : Proximité des R et L estimateurs correspondants.

En ce qui concerne les R-estimateurs, HODGES a seulement établi heuristique-ment que (voir Chapitre V, R-estimateurs) :

$$\sqrt{n}(R_n - L_n) \xrightarrow{P} 0$$

I.2.4. Comparaison d'estimateurs asymptotiquement efficaces sous F.

Pour F donné, et sous des hypothèses qui ont été détaillées dans les Chapitres IV et X, le M-estimateur asymptotiquement efficace est défini par la fonction $\psi_f = -\frac{f'}{f}$; les L et R estimateurs correspondants qui sont donc aussi asymptotiquement efficaces sous F , sont définis par :

$$h_f(t) = -\frac{1}{I(F)} \left(\frac{f'}{f}\right)' \circ F^{-1}(t) \quad t \in [0, 1]$$

$$\text{et} \quad J_f(t) = -\frac{f'}{f} \circ F^{-1}(t) \quad t \in [0, 1]$$

Nous supposons pour cela que F vérifie les hypothèses suivantes : $F \in \mathcal{H}(F)$, f est continûment différentiable sur \mathbb{R} , $\psi_f = -\frac{f'}{f}$ a une dérivée continue sauf en un nombre fini de points, $\psi_f \in \mathcal{H}(M, F)$.

Remarquons que ces conditions excluent le cas où F est normale ce qui justifie le paragraphe I.2.5.

$$\psi_f \in \mathcal{H}(M, F) \Leftrightarrow h_f \in \mathcal{H}(L) \Leftrightarrow J_f \in \mathcal{H}(R, F)$$

Notons $M(F)$, $L(F)$, $R(F)$ les estimateurs définis par ψ_f , h_f , et J_f respectivement

Ils sont donc asymptotiquement efficaces sous F , c'est-à-dire de variance asymptotique minimum égale à $\frac{1}{I(F)}$ et asymptotiquement normaux sous G pourvu que $G \in \mathcal{H}(G)$ - G symétrique, de densité $g > 0$ sur \mathbb{R} et d'information finie - et que g soit continue. En effet, il nous faut vérifier que

$$\int_{-\infty}^{+\infty} \psi_f'(x) dG(x) \neq 0 \quad \text{et que} \quad \int_0^1 J_f'(t) \cdot g \circ G^{-1}(t) dt \neq 0.$$

Il n'est pas nécessaire ici de vérifier que $J_f' \cdot g \circ G^{-1}$ est à variation bornée car cette hypothèse n'est pas nécessaire pour avoir la normalité asymptotique sous G du R -estimateur défini par J_f . Nous l'avons introduite dans $\mathcal{H}(R, G)$ pour pouvoir définir des correspondances bijectives entre les 3 ensembles de fonctions. Par hypothèse, ψ_f' est positive, non identiquement nulle, à support compact et continue sauf en un nombre fini de points. On peut donc trouver un intervalle $[\alpha, \beta]$ sur lequel ψ_f' soit continue et strictement positive.

$+\infty > \int_{\mathbb{R}} \psi_f'(x) dG(x) \geq \int_{\alpha}^{\beta} \psi_f'(x) dG(x) > 0$ puisque G a pour support \mathbb{R} tout entier.

$$D'autre part, \mathcal{J} = \int_0^1 J_f'(t) \cdot g \circ G^{-1}(t) dt = \int_0^1 (\psi_f \circ F^{-1})'(t) \cdot g \circ G^{-1}(t) dt$$

$$J_f'(t) \cdot g \circ G^{-1}(t) = \psi_f' \circ F^{-1}(t) \frac{g \circ G^{-1}(t)}{f \circ F^{-1}(t)}, J_f' \cdot g \circ G^{-1} \text{ est positive, non}$$

identiquement nulle, continue sauf en un nombre fini de points et à support compact inclus dans $]0, 1[$. Il en résulte que $0 < \mathcal{J} < +\infty$. On peut donc écrire les variances asymptotiques sous G des 3 estimateurs :

$$\sigma_G^2(M(F)) = \frac{\int_{\mathbb{R}} \psi_f^2(x) g(x) dx}{\left(\int_{\mathbb{R}} \psi_f'(x) g(x) dx\right)^2} \quad \sigma_G^2(R(F)) = \frac{\int_0^1 (\psi_f \circ F^{-1}(t))^2 dt}{\left[\int_{\mathbb{R}} (\psi_f \circ F^{-1})' \circ G(x) g^2(x) dx\right]^2}$$

$$\text{et } \sigma_G^2(L(F)) = \int_0^1 K^2(t) dt \quad \text{avec} \quad dK(t) = \frac{1}{I(F)} \psi_f' \circ F^{-1}(t) dG^{-1}(t)$$

$$\text{et } \int_0^1 K(t) dt = 0.$$

I.2.4.a. Comparaison des R et L estimateurs.Théorème :

Si F vérifie les conditions énoncées ci-dessus, $\sigma_G^2(R(F)) \leq \sigma_G^2(L(F))$ pour toutes les distributions de probabilité G, de densité $g > 0$ sur R, continue, symétrique, et d'information finie.

Démonstration : (cf. F.W. SCHOLZ [81])

$$\sigma_G^2(R(F)) = \frac{\int_R \psi_f(x)^2 f(x) dx}{\left[\int_R \frac{\psi'_f \circ F^{-1} \circ G(x)}{f \circ F^{-1} \circ G} g^2(x) dx \right]^2} = \frac{I(F)}{\left[\int_R I(F) \frac{h_f \circ G(x)}{f \circ F^{-1} \circ G(x)} g^2(x) dx \right]^2}$$

$$\Rightarrow \sigma_G^2(R(F)) = \frac{1}{I(F)} \frac{1}{\left[\int_R \frac{g(x)}{f \circ F^{-1} \circ G(x)} \cdot h_f \circ G(x) g(x) dx \right]^2}$$

Or $x \rightarrow h_f \circ G(x) \cdot g(x)$ est une densité de probabilité puisque $h_f \geq 0$ et

$$\int_R h_f \cdot g(x) g(x) dx = \int_0^1 h_f(t) dt = 1, \quad \frac{g}{f \circ F^{-1} \circ G}$$

est strictement positive par hypothèse on peut donc appliquer l'inégalité de JENSEN à la fonction convexe $\varphi : \varphi(x) = \frac{1}{x}$ pour $x > 0$.

$$\begin{aligned} \text{Soit } \frac{1}{\int_R \frac{g(x)}{f \circ F^{-1} \circ G(x)} h_f \circ G(x) g(x) dx} &\leq \int_R \frac{1}{\frac{g(x)}{f \circ F^{-1} \circ G(x)}} h_f \circ G(x) g(x) dx \\ &= \int_R f \circ F^{-1} \circ G(x) h_f \circ G(x) dx \end{aligned}$$

$$\int_{\mathbb{R}} f \circ F^{-1} \circ G(x) h_f \circ G(x) dx = \left[\left(\int_0^x h_f \circ G(y) dy \right) f \circ F^{-1} \circ G(x) \right]_{-\infty}^{+\infty} \\ - \int_{-\infty}^{+\infty} \left(\int_0^x h_f \circ G(y) dy \right) \cdot \frac{f' \circ F^{-1} \circ G(x)}{f \circ F^{-1} \circ G(x)} g(x) dx$$

$$\lim_{x \rightarrow +\infty} f \circ F^{-1} \circ G(x) = \lim_{x \rightarrow +\infty} f(x) = 0 \quad \text{puisque } \left(-\frac{f'}{f}\right)' \text{ est à support}$$

compact. De même $\lim_{x \rightarrow -\infty} f \circ F^{-1} \circ G(x) = 0$. Comme $\int_0^x h_f \circ G(y) dy$ est bornée puisque h_f est à support compact, on en déduit que

$$\left[\int_0^x h_f \circ G(y) dy \cdot f \circ F^{-1} \circ G(x) \right]_{-\infty}^{+\infty} = 0$$

$$\text{Posons } \int_0^x h_f(G(y)) dy = K_f(G(x)) = \int_{1/2}^{G(x)} h_f(t) dG^{-1}(t)$$

$$\text{On a : } \frac{1}{\int_{-\infty}^{+\infty} \frac{g(x)}{f \circ F^{-1} \circ G(x)} \cdot h_f \circ G(x) dx} \leq \int_{-\infty}^{+\infty} K_f \circ G(x) \left(-\frac{f' \circ F^{-1} \circ G(x)}{f \circ F^{-1} \circ G(x)}\right) g(x) dx \\ = \int_0^1 K_f(t) \cdot \psi_f \circ F^{-1}(t) dt$$

$$\text{D'où } \sigma_G^2(R(F)) \leq \frac{1}{I(F)} \left(\int_0^1 K_f(t) \psi_f \circ F^{-1}(t) dt \right)^2$$

$$\leq \frac{1}{I(F)} \int_0^1 K_f(t)^2 dt \int_0^1 (\psi_f \circ F^{-1}(t))^2 dt = \sigma_G^2(L(F))$$

C.Q.F.D.

I.2.4.b) Comparaison des R et M estimateurs.

Il n'est pas possible de démontrer que l'un des estimateurs est uniformément meilleur que l'autre.

Soit G une loi symétrique, fortement unimodale, de densité g positive sur \mathbb{R} et deux fois continûment dérivable ; $(-\log g)$ est convexe et $(-\log g)''$ est à variation bornée sur les compacts. Soit ϵ fixé dans $]0, 1[$.

Désignons par F_0 la loi de densité :

$$f_0(x) = \begin{cases} (1 - \epsilon) g(x) & \text{si } |x| \leq x_0 \\ (1 - \epsilon) g(x_0) e^{-k(|x| - x_0)} & \text{si } |x| \geq x_0 \end{cases}$$

où k et x_0 sont déterminés par :

$$\begin{cases} \left(\left| \frac{g'(x_0)}{g(x_0)} \right| < k \text{ sur }]-x_0, x_0[, - \frac{g'(x_0)}{g(x_0)} = k \right. \\ \left. \int_{\mathbf{R}} f_0(t) dt = 1 \right. \end{cases}$$

Alors $\psi_0 = - \frac{f'_0}{f_0} = \max(-k, \min(-\frac{g'}{g}, k))$ définit un M-estimateur efficace en F_0 , et par suite $J_0 = \psi_0 [F_0^{-1}]$ définit aussi un R-estimateur efficace en F_0 (voir I.2.1).

Nous allons comparer les variances asymptotiques de ces deux estimateurs sous deux lois F qui admettent une densité continue sauf en deux points ; nous pourrions modifier légèrement ces densités pour les rendre continues, ce qui permettrait d'avoir $\psi_0 \in \mathcal{H}(M, F)$ et $J_0 \in \mathcal{H}(R, F)$ mais ce n'est pas nécessaire : les formules de variances asymptotiques s'appliquent en utilisant les conditions du Chapitre V : R-estimateurs.

$$\begin{aligned} \sigma^2(\psi_0, F) &= \frac{\int \psi_0^2(x) f(x) dx}{\left(\int \psi'_0(x) f(x) dx \right)^2} = \frac{2k^2 F(-x_0) + \int_{-x_0}^{x_0} \psi_0^2(x) f(x) dx}{\left(\int_{-x_0}^{x_0} \psi'_0(x) f(x) dx \right)^2} \\ \sigma^2(J_0, F) &= \frac{\int_0^1 J_0^2(t) dt}{\left(\int_{-\infty}^{\infty} J_0[F(x)]' dF(x) \right)^2} = \frac{\int_0^1 \psi_0[F_0^{-1}(t)]^2 dt}{\left\{ \int_{-\infty}^{\infty} \frac{\psi'_0[F_0^{-1}(F(x))]}{f_0[F_0^{-1}(F(x))]} f^2(x) dx \right\}^2} \\ &= \frac{2k^2 F_0(-x_0) + \int_{-x_0}^{x_0} \psi_0^2(x) f_0(x) dx}{\left\{ \int_{-x_0}^{x_0} \psi'_0(x) \cdot f[F^{-1}(F_0(x))] dx \right\}^2} \end{aligned}$$

Exemple 1 :

$$\text{Soit } f(x) = \begin{cases} (1 - \epsilon) g(x) & \text{si } |x| \geq \alpha \\ \frac{1 - 2(1 - \epsilon) G(-\alpha)}{2\alpha} & \text{si } |x| < \alpha \end{cases}$$

α est choisi suffisamment petit pour que $\frac{1 - 2(1 - \epsilon) G(-\alpha)}{2\alpha} > (1 - \epsilon) g(0)$ de sorte que $f \in C$.

Comparons d'abord les numérateurs des variances asymptotiques :

$$\int_{-x_0}^{x_0} \psi_0^2(x) f(x) dx \xrightarrow{\alpha \rightarrow 0} \int_{-x_0}^{x_0} \psi_0^2(x) \cdot (1 - \epsilon) g(x) dx = \int_{-x_0}^{x_0} \psi_0^2(x) f_0(x) dx$$

car $\psi_0(0) = 0$.

De $F(-x_0) = F_0(-x_0) - \frac{\epsilon}{2}$, nous tirons immédiatement :

$$\int_{\mathbb{R}} \psi_0^2(x) f(x) dx \longrightarrow \int_{\mathbb{R}} \psi_0^2(x) f_0(x) dx - \epsilon \cdot k^2$$

Comparons maintenant les dénominateurs :

$$\begin{aligned} \int_{-x_0}^{x_0} \psi_0'(x) \cdot f[F^{-1}(F_0(x))] dx &= \int_{-x_0}^{F_0^{-1}[F(-\alpha)]} \psi_0'(x) \cdot (1 - \epsilon) \cdot g[F^{-1}(F_0(x))] dx \\ &+ \int_{F_0^{-1}[F(-\alpha)]}^{F_0^{-1}[F(\alpha)]} \psi_0'(x) \cdot \frac{1 - 2G(-\alpha)(1 - \epsilon)}{2\alpha} \cdot dx \\ &+ \int_{F_0^{-1}[F(\alpha)]}^{x_0} \psi_0'(x) (1 - \epsilon) \cdot g[F^{-1}(F_0(x))] dx \end{aligned}$$

La somme du 1er terme et du 3ème terme tend vers

$$\int_{-x_0}^{x_0} \psi_0'(x) (1 - \epsilon) g[F^{-1}(F_0(x))] dx \text{ quand } \alpha \text{ tend vers } 0.$$

Le 2ème terme vaut :

$$\frac{1 - 2 G(-\alpha)(1-\epsilon)}{2\alpha} \cdot 2 \psi'_0 [F_0^{-1}(F(\alpha))] \text{ qui est équivalent à } \frac{\epsilon}{2\alpha} \cdot 2 [\psi_0(o) + \alpha \cdot \frac{f(o) \cdot \psi'_0(o)}{f_0(o)}], \text{ soit } \frac{\epsilon}{2\alpha} \cdot \frac{\psi'_0(o)}{(1-\epsilon)g(o)} \text{ donc}$$

$$\int_{-x_0}^{x_0} \psi'_0(x) \cdot f[F_0^{-1}(F_0(x))] dx \xrightarrow{\alpha \rightarrow 0} \infty$$

$$\int_{-x_0}^{x_0} \psi'_0(x) f(x) dx = \int_{-x_0}^{-\alpha} \psi'_0(x) \cdot (1-\epsilon)g(x) dx + \int_{-\alpha}^{\alpha} \psi'_0(x) \cdot \frac{1-2G(-\alpha)(1-\epsilon)}{2\alpha} \cdot dx$$

$$+ \int_{\alpha}^{x_0} \psi'_0(x) \cdot (1-\epsilon)g(x) dx \xrightarrow{\alpha \rightarrow 0} \int_{-x_0}^{x_0} \psi'_0(x) (1-\epsilon)g(x) dx$$

$$+ \psi'_0(0) [1 - 2 G(o)(1-\epsilon)] = (1-\epsilon) \int_{-x_0}^{x_0} \psi'_0(x) \cdot g(x) dx + \epsilon \cdot \psi'_0(o)$$

En conclusion, $\sigma^2(J_\alpha, F) \xrightarrow{\alpha \rightarrow 0} 0$ alors que $\sigma^2(\psi_0, F)$ tend vers une limite non nulle. Ce qui montre que pour α suffisamment petit : $\sigma^2(J_\alpha, F) < \sigma^2(\psi_0, F)$.

Exemple 2 :

i) Si $\frac{1 - (1-\epsilon)G(-x_0)}{2x_0} > (1-\epsilon)g(o)$, nous pouvons choisir l'élément de C suivant :

$$f(x) = \begin{cases} (1-\epsilon)g(x) & \text{si } |x| \geq x_0 \\ \frac{1 - (1-\epsilon)G(-x_0)}{2x_0} & \text{si } |x| < x_0 \end{cases}$$

Pour tout x compris entre $-x_0$ et 0, on a : $F_0(x) > F(x)$ et par suite :

$$-x_0 < x < F_0^{-1}(F_0(x)) \leq 0 \text{ donc } f(x) = f[F_0^{-1}(F_0(x))]$$

sur l'intervalle $]-x_0, x_0[$ par symétrie : les dénominateurs des variances asymptotiques sont égaux.

ii) Si $\frac{1 - (1-\epsilon)G(-x_0)}{2x_0} \leq (1-\epsilon)g(o)$, il existe x_1 tel que $f(x)$ défini par

$$\left. \begin{array}{ll} ((1 - \epsilon) g(x_1)) & \text{si } x_1 \leq |x| < x_0 \\ ((1 - \epsilon) g(x)) & \text{si } |x| < x_1 \text{ ou } |x| \geq x_0 \end{array} \right\} \text{ soit une densité ;}$$

$[x_1$ est l'unique solution de : $\frac{\epsilon}{2} + (1 - \epsilon) [G(x_0) - G(x_1)] = (x_0 - x_1) (1 - \epsilon) g(x_1)]$.

Pour tout x compris entre $-x_0$ et 0, on a : $F_0(x) \geq F(x)$ donc $F^{-1}(F_0(x)) \geq x$.
avec de plus $F^{-1}(F_0(x)) = x$ pour tout x compris entre $-x_1$ et 0. On en déduit :

$$-x_0 < x \leq F^{-1}(F_0(x)) \leq F^{-1}(F_0(-x_1)) = -x_1$$

ce qui permet d'avoir encore $f(x) = f[F^{-1}(F_0(x))]$ pour tout x compris entre $-x_0$ et x_0 et donc l'égalité des dénominateurs des variances asymptotiques.

Comparons les numérateurs dans les deux cas i) et ii) :

$$\begin{aligned} 2k^2 F(-x_0) + \int_{-x_0}^{x_0} \psi_0^2(x) f(x) dx &= 2k^2 [F_0(-x_0) - \frac{\epsilon}{2}] + \int_{-x_0}^{x_0} \psi_0^2(x) [f(x) - f_0(x)] dx \\ &+ \int_{-x_0}^{x_0} \psi_0^2(x) f_0(x) dx = 2k^2 F_0(-x_0) + \int_{-x_0}^{x_0} \psi_0^2(x) f_0(x) dx \\ &- \epsilon \cdot k^2 + \int_{-x_0}^{x_0} \psi_0^2(x) [f(x) - f_0(x)] dx \end{aligned}$$

Par hypothèse, on a $|\psi_0(x)| = |-\frac{g'(x)}{g(x)}| < k$ pour tout x compris entre $-x_0$ et x_0 . Ce qui permet de conclure :

$$\begin{aligned} \int_{-x_0}^{x_0} \psi_0^2(x) [f(x) - f_0(x)] dx &< \int_{-x_0}^{x_0} k^2 [f(x) - f_0(x)] dx = k^2 \cdot \epsilon \\ \sigma^2(\psi_0, F) &< \sigma^2(J_0, F) \end{aligned}$$

1.2.5. Comparaison d'estimateurs asymptotiquement efficaces pour la loi normale.

Soit ϕ la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$ et φ sa densité.

$$\psi(x) = -\frac{\varphi'(x)}{\varphi(x)} = x$$

$$h(t) = \psi' \circ \phi^{-1}(t) = 1 \text{ pour tout } t$$

$$J(t) = \psi \circ \phi^{-1}(t) = \phi^{-1}(t)$$

Ces trois fonctions ne vérifient pas les hypothèses que nous avons imposées.

Cependant, le M-estimateur défini par $\psi(x) = x$ et le L-estimateur défini par $h(t) = 1$ sont égaux à $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$. Donc, si G est une distribution de probabilité symétrique, de densité g , admettant un moment d'ordre 2, on sait que $\sqrt{n} \cdot \bar{X}_n$ converge en loi vers $\mathcal{N}(0, \sigma_G^2)$ où $\sigma_G^2 = \int_{\mathbb{R}} x^2 g(x) dx - \left(\int_{\mathbb{R}} x g(x) dx \right)^2$. Ainsi : $\sigma_G^2(\bar{X}_n) = \sigma_G^2$.

D'autre part, le R-estimateur R_n défini par $J(t) = \phi^{-1}(t)$ est le R-estimateur construit à partir du test de rang dit "de Van Der Waerden" (cf. Chapitre V) défini par les scores $b_n(i) = \phi^{-1}\left(\frac{1}{2} + \frac{i}{2n}\right)$, test de même puissance asymptotique sous ϕ que le test de rang dit "de Fisher-Yates" défini par les scores $b'_n(i) = E(|V|^{(i)})$, $|V|^{(i)}$ étant la i^{e} statistique d'ordre de l'échantillon $(|V_1|, \dots, |V_n|)$ où les V_i sont indépendantes et de loi $\mathcal{N}(0, 1)$.

On sait que si G est d'information finie $I(G) < +\infty$, alors $\sqrt{n} R_n$ converge en loi vers $\mathcal{N}(0, \sigma_G^2(R))$ avec

$$\sigma_G^2(R) = \frac{\int_0^1 \phi^{-1}(t)^2 dt}{\left[\int_{-\infty}^{+\infty} (\phi^{-1})' \circ G(x) g^2(x) dx \right]^2} = \frac{1}{I_{1G}^2}$$

avec $I_{1G} = \int_{-\infty}^{+\infty} (\phi^{-1})' \circ G(x) g^2(x) dx$; pourvu que $I_{1G} \neq 0$

(cf. Chapitre V - Théorème de convergence en loi vers normale avec faibles hypothèses sur J).

Si on impose en plus à g d'être continue comme dans le paragraphe précédent, ceci est réalisé.

Notons $E_{c_1, t}(G)$, l'efficacité relative asymptotique sous G , du test c_1 de Fisher-Yates, par rapport au test t de Student. D'après [17], $E_{c_1, t}(G) = I_{1G}^2 \cdot \sigma_G^2$,

$E_{c_1, t}(G) \geq 1$ pour toute loi G , admettant une densité et un moment d'ordre 2, et

$E_{c_1, t}(G) = 1$ si et seulement si G est normale. Comme

$$E_{c_1, t}(G) = \frac{\sigma_G^2(\bar{X}_n)}{\sigma_G^2(R_n)},$$

on peut conclure :

Propriété : Si G est symétrique, d'information finie, de densité g continue et admettant un moment d'ordre 2, alors :

$$\sigma_G^2(R_n) \leq \sigma_G^2(\bar{X}_n)$$

l'égalité n'ayant lieu que si G est normale.

II - UN PROBLÈME DE MINIMAX.

Soit C , $C = C(\epsilon, G)$ l'ensemble des distributions de probabilité F de la forme $F = (1 - \epsilon) G + \epsilon H$ où ϵ est fixé dans $]0, 1[$, G est fixée, symétrique et fortement unimodale et H est une loi symétrique quelconque.

On suppose que G a une densité g , $g > 0$ sur \mathbb{R} , deux fois continûment dérivable, telle que $(-\log g)$ est convexe et $(-\log g)''$ à variation bornée sur les compacts.

Rappelons ici un résultat d'HUBER (cf. aussi Chapitre X) :

Il existe $F_0 \in C$ et $\psi_0 \in \mathcal{H}(M, G)$ tels que

$$\sup_{F \in C} \sigma^2(\psi_0, F) = \sigma^2(\psi_0, F_0) = \inf_{\psi \in \mathcal{H}(M, G)} \sigma^2(\psi, F_0)$$

où $\sigma^2(\psi, F)$ désigne la variance asymptotique du M-estimateur défini par ψ lorsque F est la distribution sous-jacente.

Remarquons que si $\psi \in \mathcal{H}(M, G)$, alors pour tout $F \in C$, $\psi \in \mathcal{H}(M, F)$. La loi la plus défavorable pour l'estimateur minimax ψ_0 a pour densité f_0 :

$$f_0(t) = \begin{cases} (1 - \epsilon) g(t_0) e^{-k(|t| - t_0)} & \text{si } |t| \geq t_0 \\ (1 - \epsilon) g(t) & \text{si } |t| < t_0 \end{cases}$$

$$\psi_0 = - \frac{f'_0}{f_0} \quad \sigma^2(\psi_0, F_0) = \frac{1}{I(F_0)}$$

k et t_0 sont déterminés par :

$$(1) \quad \left| \frac{g'}{g} \right| \leq k \text{ sur } [-t_0, +t_0], \text{ l'égalité ayant lieu en } +t_0 \text{ et } -t_0$$

$$(2) \quad \int_{-\infty}^{+\infty} f_0(t) dt = 1$$

$$(\text{Soit } \frac{1}{1 - \epsilon} = \int_{-t_0}^{t_0} g(t) dt + 2 \frac{g(t_0)}{k})$$

$$\psi_0(x) = \begin{cases} -k & x \leq -t_0 \\ -\frac{g'}{g} & |x| < t_0 \\ k & x \geq t_0 \end{cases}$$

Et les hypothèses sur G entraînent que $\psi_0 \in \mathcal{H}(M, G)$.

Théorème :

Soient $\tilde{\mathcal{C}}$ l'ensemble des distributions $F \in \mathcal{C}$ admettant une densité f (f est nécessairement strictement positive) et $\tilde{\mathcal{C}}$ l'ensemble des distributions $F \in \tilde{\mathcal{C}}$ d'information finie, de densité f continue.

Les estimateurs définis par $h_0(t) = \psi_0'(F_0^{-1}(t))$ et $J_0(t) = \psi_0[F_0^{-1}(t)]$ sont solutions du problème de minimax pour $F \in \tilde{\mathcal{C}}$ et $h \in \mathcal{H}(L)$, et pour $F \in \tilde{\mathcal{C}}$ et $J \in \mathcal{H}(R, F_0)$.

Démonstration : On vérifie immédiatement que $f_0(x) > 0$ pour tout $x \in \mathbb{R}$ et que $I(F_0) < +\infty$. D'autre part, ψ_0 vérifie les hypothèses $\mathcal{H}(M, F_0)$, donc h_0 et J_0 définissent respectivement un L et un R-estimateur ($L_{(h_0)}$ et $R_{(J_0)}$) ayant même variance asymptotique sous F_0 :

$$\sigma^2_{(\psi_0, F_0)} = \sigma^2_{(J_0, F_0)} = \sigma^2_{(h_0, F_0)}$$

Comme $M_{(\psi_0)}$ est asymptotiquement efficace en F_0 , $L_{(h_0)}$ et $R_{(J_0)}$ le sont aussi. Ainsi :

$$\sigma^2_{(J_0, F_0)} = \inf_{J \in \mathcal{H}(R, F_0)} \sigma^2(J, F_0)$$

et
$$\sigma^2_{(h_0, F_0)} = \inf_{h \in \mathcal{H}(L)} \sigma^2(h, F_0)$$

$$\begin{aligned} \text{Il reste donc à montrer que } \sigma^2_{(h_0, F)} &\leq \sigma^2_{(h_0, F_0)} \quad \forall F \in \tilde{\mathcal{C}} \text{ et} \\ \text{et } \sigma^2_{(J_0, F)} &\leq \sigma^2_{(J_0, F_0)} \quad \forall F \in \tilde{\mathcal{C}} \end{aligned}$$

après avoir vérifié que $L_{(h_0)}$ est asymptotiquement normal sous F , $F \in \tilde{\mathcal{C}}$ de variance asymptotique $\sigma^2_{(h_0, F)}$ et que $R_{(J_0)}$ est asymptotiquement normal sous F ,

$F \in \tilde{\mathcal{C}}$, de variance asymptotique $\sigma^2(J_0, F)$.

Lemme : $\forall F \in \tilde{\mathcal{C}}$, $\forall t \in [\frac{1}{2}, F_0(t_0)]$ $f(F^{-1}(t)) \geq f_0(F_0^{-1}(t))$

Démonstration : Soit $F \in \tilde{\mathcal{C}}$, $F(0) = F_0(0) = \frac{1}{2}$ puisque les distributions sont symétriques.

$$F = (1 - \epsilon) G + \epsilon H$$

$$F_0 = (1 - \epsilon) G + \epsilon H_0.$$

Pour $0 \leq x \leq t_0$ $f_0(x) = (1 - \epsilon) g(x) \leq f(x)$. Donc $F_0(x) \leq F(x)$.

Pour $\frac{1}{2} \leq t \leq F_0(t_0) \leq F(t_0)$, $0 \leq F^{-1}(t) \leq t_0 \Rightarrow F_0[F^{-1}(t)] \leq F[F^{-1}(t)] = t$

D'où $F^{-1}(t) \leq F_0^{-1}(t) \leq F_0^{-1}(F_0(t_0)) \leq t_0$.

Finalement, $\frac{1}{2} \leq t \leq F_0(t_0) \Rightarrow 0 \leq F^{-1}(t) \leq F_0^{-1}(t) \leq t_0$.

Comme g est supposée unimodale, elle est décroissante sur $[0, t_0]$; f_0 l'est donc aussi, ce qui implique $f(F^{-1}(t)) \geq f_0(F^{-1}(t)) \geq f_0(F_0^{-1}(t))$ pour tout $t \in [\frac{1}{2}, F_0(t_0)]$.

Démonstration du théorème (suite) :

$-h_0(t) = \psi_0[F_0^{-1}(t)] \geq 0$ vérifie les hypothèses $\mathcal{H}(L)$ puisque $\psi_0 \in \mathcal{H}(F_0)$ et $F_0 \in \mathcal{H}(F)$. Comme g est de classe C^2 , h_0 est continue. Donc, d'après un théorème de HUBER : [46], $L(h_0)$ est asymptotiquement normal, sans conditions sur F , donc en particulier pour $F \in \tilde{\mathcal{C}}$, de variance asymptotique $\sigma^2(h_0, F)$.

Pour $\frac{1}{2} \leq u \leq F_0(t_0)$, on a d'après le lemme $\frac{h_0(u)}{f(F^{-1}(u))} \leq \frac{h_0(u)}{f_0(F_0^{-1}(u))}$.

D'où, en intégrant, $\frac{1}{2} \leq t \leq F_0(t_0)$

$$K(t) = \int_{1/2}^t \frac{h_0(u) du}{f(F^{-1}(u))} \leq K_0(t) = \int_{1/2}^t \frac{h_0(u) du}{f_0(F_0^{-1}(u))}$$

Si $u \geq F_0(t_0)$, $h_0(u) = \psi'_0(F_0^{-1}(u)) = 0$. Donc, pour $t \geq F_0(t_0)$, on a aussi : $K(t) \leq K_0(t)$. Soit pour tout $t \in [\frac{1}{2}, 1]$, $0 \leq K(t) \leq K_0(t)$. Comme $K(1-t) = -K(t)$ et $K_0(1-t) = -K_0(t)$, on obtient $K^2(t) \leq K_0^2(t)$ pour tout $t \in [0, 1]$. D'où, $\sigma^2(h_0, F) \leq \sigma^2(h_0, F_0)$ si $F \in \tilde{\mathcal{C}}$.

- Pour $F \in \tilde{\mathcal{C}}$, $I(F) < +\infty$, on sait alors (Cf. Chapitre V) que le R-estimateur défini par $J_0(t) = \psi_0[F_0^{-1}(t)]$ est asymptotiquement normal sous F , de variance asymptotique $\sigma^2(J_0, F)$ pourvu que $\int_R (J_0 \circ F)'(x) \cdot f(x) \cdot dx \neq 0$, ce qui est réalisé puisque f continue (cf. paragraphe I.2.4).

$$\sigma^2(J_0, F) = \frac{\int_0^1 J_0^2(t) dt}{[\int_{-\infty}^{+\infty} [J_0(F(x))]' f(x) dx]^2}$$

$$\text{et} \quad \sigma^2(J_0, F_0) = \frac{\int_0^1 J_0^2(t) dt}{[\int_{-\infty}^{+\infty} [J_0(F_0(x))]' f_0(x) dx]^2}$$

Il suffit de comparer les dénominateurs.

$$\text{Soient } I = \int_{-\infty}^{+\infty} (J_0[F(x)])' f(x) dx \quad \text{et} \quad I_0 = \int_{-\infty}^{+\infty} (J_0[F_0(x)])' f_0(x) dx$$

I et I_0 sont positifs.

$$I = \int_0^1 J'_0(t) f(F^{-1}(t)) dt = 2 \int_{1/2}^1 J'_0(t) \cdot f(F^{-1}(t)) dt$$

$$\text{Or } J'_0(t) = \frac{\psi'_0(F_0^{-1}(t))}{f_0(F_0^{-1}(t))} \quad \text{est nulle pour } t \geq F_0(t_0).$$

$$\text{Donc } I = 2 \int_{1/2}^{F_0(t_0)} J'_0(t) \cdot f(F^{-1}(t)) dt. \quad \text{De même}$$

$$I_0 = 2 \int_{1/2}^{F_0(t_0)} J'_0(t) f_0(F_0^{-1}(t)) dt$$

$$\forall t \in [\frac{1}{2}, F_0(t_0)] \quad J'_0(t) \cdot f(F^{-1}(t)) \geq J'_0(t) f_0(F_0^{-1}(t))$$

puisque $J'_0 \geq 0$. Donc $I \geq I_0$. Soit, enfin, $\sigma^2_{(J_0, F)} \leq \sigma^2_{(J_0, F_0)}$ si $F \in \tilde{\mathcal{C}}$.

Remarque : Il n'est pas nécessaire de supposer f continue sur \mathbb{R} , tout entier, mais seulement continue sauf en un nombre fini de points.

Valentine COSTA
Université de Paris-Sud
Mathématiques
91405 - ORSAY

Jean DESHAYES
Université de Paris VII
Mathématiques
2, Place Jussieu
75221 PARIS CEDEX 05.

CHAPITRE XIV

COURBE D'INFLUENCE ET SENSIBILITÉ

par

Claude DENIAU – Georges OPPENHEIM – Claude VIANO

Un article de TUKEY [92] porte le titre suivant : "Which part of the sample contains the information". Dans la pratique statistique cette question est courante bien que les notions utiles pour y répondre fassent encore cruellement défaut.

On dispose cependant de quelques outils d'appréciation :

- . de ce que l'on peut appeler l'influence d'une observation sur la valeur d'une statistique (effet par exemple de l'ajout d'une observation à un échantillon) ;
- . de l'accroissement de la valeur d'une fonctionnelle dû à la contamination d'une loi par une autre .

Les notions centrales de ce chapitre sont les courbes de sensibilité et d'influence. Cette dernière est étroitement liée à la dérivée au sens de VOLTERRA ou de Von MISES d'une fonctionnelle. Ces outils sont à mettre en relation avec l'étude de la robustesse au sens de HAMPEL, développée au Chapitre IX .

Comme on le verra, seuls certains des problèmes soulevés sont résolus et dans tous les cas nous avons gardé en vue l'étude des techniques d'estimation de paramètres unidimensionnels de position.

1 - SENSIBILITÉ D'UNE STATISTIQUE À UNE MODIFICATION DE L'ÉCHANTILLON.

Soit (T_n) une suite de statistiques et $\omega = (x_1, \dots, x_n) \in \mathbb{R}^n$ un échantillon de taille n .

L'influence de l'adjonction d'une observation $x \in \mathbb{R}$ se définit et s'étudie par la courbe de sensibilité :

$$\psi_n(x, \omega) = T_{n+1}(x, \omega) - T_n(\omega)$$

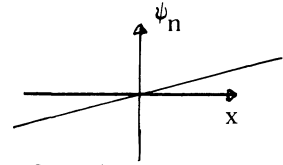
où l'on note $(x, \omega) = (x, x_1, x_2, \dots, x_n)$.

Dans les trois exemples suivants on suppose que $T_n(\omega) = 0$:

$$\psi_n(x, \omega) = T_{n+1}(x, \omega) .$$

Exemple 1 : Moyenne

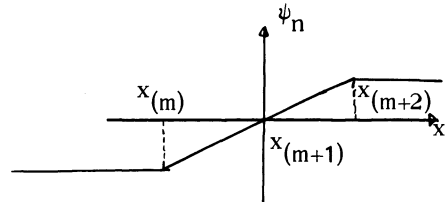
$$T_{n+1}(x, \omega) = (x + \sum_{i=1}^n x_i) / (n+1) = x / (n+1) .$$



Exemple 2 : Médiane (on note $x_{(p)}$ le quantile d'ordre p) . $n = 2m + 1$.

$$T_n(\omega) = 0 = x_{(m+1)} .$$

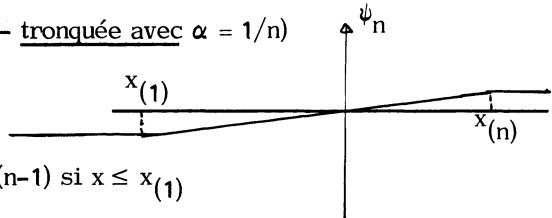
$$T_{n+1}(x, \omega) = \begin{cases} x_{(m)}/2 & \text{si } x \leq x_{(m)} \\ x/2 & \text{si } x_{(m)} \leq x \leq x_{(m+2)} \\ x_{(m+2)}/2 & \text{si } x_{(m+2)} \leq x . \end{cases}$$



Exemple 3 : Moyenne $1/n$ - tronquée (α - tronquée avec $\alpha = 1/n$)

$$T_n(\omega) = 0 = \sum_{i=2}^{n-1} x_{(i)} / (n-2) .$$

$$T_{n+1}(x, \omega) = \begin{cases} \sum_{i=1}^{n-1} x_{(i)} / (n-1) = x_{(1)} / (n-1) & \text{si } x \leq x_{(1)} \\ (x + \sum_{i=2}^{n-1} x_{(i)}) / (n-1) = x / (n-1) & \text{si } x_{(1)} \leq x \leq x_{(n)} \\ x_{(n)} / (n-1) & \text{si } x_{(n)} \leq x . \end{cases}$$



Commentaires sur ces fonctions.

Com. 1 : Alors que, pour la moyenne, la courbe de sensibilité n'est pas bornée, elle l'est pour les autres exemples envisagés :

L'adjonction d'une observation s'écartant d'une quantité x de la moyenne fait croître celle-ci de $x/(n+1)$, quantité pouvant être grande. Par contre, la moyenne 1/n-tronquée ne peut être modifiée de plus de $|x_{(n)} - x_{(1)}|/(n-1)$ quelle que soit la valeur de l'observation supplémentaire.

La quantité $\sup_x |\psi_n(x, \omega)|$ est appelée sensibilité de T_n pour l'échantillon ω . (Les valeurs de la sensibilité sont présentées dans le tableau 1).

Com. 2 : La forme de ψ_n fait apparaître le rôle particulier de certaines quantiles (ici, ceux qui correspondent aux discontinuités de ψ'_n) :

pour la médiane : $x_{(m)}$ et $x_{(m+2)}$ ($n = 2m+1$)
 pour la moyenne 1-tronquée : $x_{(1)}$ et $x_{(n)}$.

Com. 3 : Déplacement d'une observation.

La différence $\psi_n(x, \omega) - \psi_n(y, \omega)$ compare les effets de l'adjonction d'une observation en x et en y . Plutôt que cette fonction de x, y, ω on utilise :

$$(1) \quad u = \sup_{x \neq y} \frac{|\psi_n(x, \omega) - \psi_n(y, \omega)|}{|x - y|} = \sup_{x \neq y} \frac{|T_{n+1}(x, \omega) - T_{n+1}(y, \omega)|}{|x - y|} .$$

Ces quantités peuvent servir d'indice pour l'étude de l'effet du déplacement d'une observation de x en y : pour la médiane par exemple, la modification est la plus grande lorsqu'on déplace l'observation sur l'intervalle $]x_{(m)}, x_{(m+2)}[$.

Statistique	Sensibilité	u
Moyenne	∞	1
Médiane ($n = 2m + 1$)	$(x_{(m+2)} - x_{(m)}) / 2$	1/2
Moyenne 1-tronquée	$(x_{(n)} - x_{(1)}) / (n-1)$	1/(n-1)

Tableau 1 : Sensibilité des statistiques et quantité u définie par (1).

En résumé, on a introduit plusieurs indices descriptifs de l'effet sur la statistique T_n de modifications simples de l'échantillon (ajout, déplacement ou suppression d'une observation).

2 - INFLUENCE SUR UNE FONCTIONNELLE D'UNE CONTAMINATION DE LA LOI.

Les indices introduits dans le paragraphe précédent dépendent souvent des valeurs des observations, de l'échantillon ω et de n . Pour les études asymptotiques et pour les statistiques restrictions de fonctionnelle, il s'avère intéressant de généraliser ces indices en introduisant, après quelques définitions, la courbe d'influence.

Différentielle directionnelle.

Un domaine $\mathcal{D} \in \mathcal{F}$ est étoilé en $F \in \mathcal{D}$ si et seulement si pour tout $G \in \mathcal{D}$ le segment $[F, G] = \{ F + \epsilon (G - F) \mid 0 \leq \epsilon \leq 1, \epsilon \in \mathbb{R} \}$ est contenu dans \mathcal{D} .

On montre que tout convexe C de \mathcal{F} est étoilé en tout point de C et que par conséquent une boule pour la métrique de PROHOROV est étoilée, par exemple, en son centre.

Soit \mathcal{D} un domaine étoilé en F et G un point de \mathcal{D} .

Une fonction $T : \mathcal{F} \rightarrow \mathbb{R}$ mesurable, est dite différentiable en $F \in \mathcal{D}$ dans la direction de $G \in \mathcal{D}$ si et si et seulement il existe $u(F, G) \in \mathbb{R}$ tel que :

$$\lim_{\epsilon \downarrow 0} \left[\frac{T(F + \epsilon(G - F)) - T(F)}{\epsilon} - u(F, G) \right] = 0.$$

Courbe d'influence.

Soit \mathcal{D} le domaine étoilé en F engendré par la famille des segments $[F, \delta_x]$, δ_x étant la mesure de DIRAC en x .

Si T est dérivable en F dans la direction δ_x la dérivée est appelée courbe d'influence de T en F et notée :

$$T'(F, x) = \lim_{\epsilon \downarrow 0} \frac{T(F + \epsilon(\delta_x - F)) - T(F)}{\epsilon}$$

L'écriture

$$\left(\frac{d}{d\epsilon} T(F + \epsilon(\delta_x - F)) \right)_{\epsilon=0} = \int T'(F, x)(\delta_x - F)(dx)$$

liant T' à la dérivée en $\epsilon = 0$ de $T(F + \epsilon(\delta_x - F))$ permet de relier la dérivée directionnelle introduite à la famille des dérivées usuelles en analyse fonctionnelle (dérivée de VOLTERRA, de Von MISES) et indique aussi que l'étude de la robustesse à venir sera dans une large mesure celle des propriétés de dérivabilité et continuité des fonctionnelles.

Exemples : Nous étudions quelques fonctionnelles dont les restrictions à \mathfrak{F}_n sont des estimateurs fréquemment rencontrés. ANDREWS [2] fournit les courbes d'influence d'un grand nombre de fonctionnelles. On supposera ici que F est absolument continue par rapport à la mesure de LEBESGUE et on notera f sa densité.

Pour les exemples suivants on peut se reporter à ANDREWS ([2], p. 34-35 et 49) Seul le calcul concernant les M-estimateurs (exemple 7) est présenté ici. Dans les autres cas les résultats sont fournis.

Exemple 4 : Espérance

$$T(F) = \int x F(dx), \text{ si } T(F) < \infty. \quad T'(F, x) = x - T(F).$$

Exemple 5 : Espérance α -tronquée ($0 < \alpha < \frac{1}{2}$)

$$T(F) = \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} F^{-1}(t) dt$$

où F^{-1} est la fonction inverse de F .

Si l'information de FISHER $I(F)$ est finie, f symétrique et unimodale, notons

$$a = F^{-1}(\alpha) \quad b = F^{-1}(1-\alpha)$$

$$T'(F, x) = \begin{cases} a/(1-2\alpha) & \text{si } x \leq a \\ x/(1-2\alpha) & \text{si } a \leq x \leq b \\ b/(1-2\alpha) & \text{si } b \leq x \end{cases}$$

Exemple 6 : Espérance α -Winzorisée ($0 < \alpha < \frac{1}{2}$)

$$T(F) = \alpha(a+b) + \int_{\alpha}^{1-\alpha} F^{-1}(t) dt$$

Si $I(F)$ est finie, f symétrique et unimodale

$$T'(F, x) = \begin{cases} a - (\alpha/f(a)) & \text{si } x < a \\ x & \text{si } a \leq x \leq b \\ b + (\alpha/f(b)) & \text{si } b < x \end{cases}$$

Exemple 7 : R-estimateurs

Avec les définitions du Chapitre IV, si $I(F)$ est finie, si f est symétrique, unimodale et admet \mathbb{R} pour support. Posons :

$$u'(x) = J'(F(x)) f(x) \quad \text{et} \quad u(x) = \int_{-\infty}^x u'(y) dy.$$

$$\text{Alors} \quad T'(F, x) = [u(x) - \int u(x) f(x) dx] / \int u'(x) f(x) dx.$$

Pour l'estimateur de HODGES-LEHMAN associé à $J(t) = t - \frac{1}{2}$ on obtient :

$$T'(F, x) = (F(x) - \frac{1}{2}) / \int f^2(x) dx.$$

Exemple 8 : L-estimateurs

Avec les définitions du Chapitre V, si $I(F)$ est finie et, si :

$$\int_0^1 J(t) dt = 1 \quad \text{alors} \quad T(F) = \int_0^1 J(t) F^{-1}(t) dt$$

et

$$T'(F, x) = \int_0^1 J(t) \frac{t - \delta_x(F^{-1}(t))}{f(F^{-1}(t))} dt.$$

Exemple 9: M-estimateurs

Avec les définitions du Chapitre III on obtient les résultats suivants :

$T(F)$ est Volterra-dérivable à l'ordre 1 en F sur le domaine étoilé \mathcal{D} engendré par les lois G et

$$\left(\frac{dT(F + \epsilon(G-F))}{d\epsilon} \right)_{\epsilon=0} = \left(\frac{dt_G(\epsilon)}{d\epsilon} \right)_{\epsilon=0} = - \frac{\int \psi(x - T(F)) G(dx)}{\lambda'_F(T(F))}$$

dès que ψ et F vérifient les conditions a et b et pour les lois G vérifiant la condition c :

a) ψ est continue et F intégrable. $I(F)$ est finie.

b) $\lambda_F(a) = \int \psi(x - a) dF(x)$ est définie et l'équation $\lambda_F(a) = 0$ admet une solution unique $T(F)$. $\lambda_F(a)$ est dérivable au voisinage de $T(F)$ et $\lambda'_F(T(F)) \neq 0$.

c) ψ est G intégrable. $\lambda_G(a)$ est continue au voisinage de $T(F)$ et $t_G(\epsilon)$ est continu en $\epsilon = 0$.

Preuve : Posons $F_\epsilon = (1 - \epsilon)F + \epsilon G$. Les conditions a et c entraînent que dans un voisinage de $\epsilon = 0$ l'équation $\lambda_{F_\epsilon}(a) = 0$ admet une solution. Donc dans ce voisinage $\lambda_{F_\epsilon}(T(F_\epsilon)) = 0 = (1 - \epsilon) \lambda_F(T(F_\epsilon)) + \epsilon \lambda_G(T(F_\epsilon))$. La dérivabilité de $\lambda_F(a)$ permet d'écrire :

$$\lambda_{F_\epsilon}(T(F_\epsilon)) = \lambda_F(T(F)) + (T(F_\epsilon) - T(F)) [\lambda'_F(T(F)) + o(T(F_\epsilon) - T(F))].$$

$$\text{Soit } \frac{t_G(\epsilon) - T(F)}{\epsilon} = - \frac{\lambda_G(T(F_\epsilon))}{(1-\epsilon) [\lambda'_F(T(F)) + o(t_G(\epsilon) - T(F))]}$$

d'où le résultat annoncé en faisant tendre ϵ vers zéro et en utilisant la continuité de $\lambda_G(a)$ et celle de $t_G(\epsilon)$.

. Courbe d'influence : On pose $G = \delta_x$ et on suppose que ψ et F vérifient les conditions a, b et c (notons qu'alors la continuité de ψ entraîne celle de $\lambda_{\delta_x}(a)$) alors :

$$T'(F, x) = - \frac{\psi(x - T(F))}{\lambda'_F(T(F))}.$$

Remarque : Les conditions a, b et c ne sont pas nécessaires. Le résultat reste valable par exemple sous les conditions :

- . $I(F)$ finie. ψ est continue monotone croissante de signe non constant.
- . $\lambda_F(a) = 0$ a une solution unique $T(F)$. Elle est dérivable au point $T(F)$ et sa dérivée y est non nulle.

Remarquons encore que le résultat est valable pour la médiane alors que la fonction ψ correspondante n'est pas continue .

Exemple 10 : La médiane

$$T(F) = F^{-1}\left(\frac{1}{2}\right).$$

Si $I(F)$ est finie et si f est symétrique et unimodale :

$$T'(F, x) = \begin{cases} -1/2 f(0) & \text{si } x < 0 \\ 0 & \text{si } x = 0 \\ 1/2 f(0) & \text{si } x > 0 \end{cases}$$

On pourra comparer les résultats des exemples 1 et 4, 2 et 10, 3 et 5.

3 - UTILISATION DE LA COURBE D'INFLUENCE.

L'existence de la dérivée de VOLTERRA ou de Von MISES d'une fonctionnelle T permet d'écrire des développements à l'ordre 1 du type :

$$T(F + \epsilon(G - F)) - T(F) = \epsilon \int T'(F, x)(G - F)(dx) + o(\epsilon)$$

$$\text{ou } T(F + \epsilon(\delta_{x_0} - F)) - T(F) = \epsilon \int T'(F, x)(\delta_{x_0} - F)(dx) + o(\epsilon)$$

$$(2) \quad = \epsilon T'(F, x_0) + o(\epsilon) .$$

(a). Une première utilisation possible de ces expressions repose sur la propriété suivante :

Si T est une fonctionnelle faiblement continue en F et si pour tout n $T_n = T \mid M_n$ est la restriction de T à \mathfrak{F}_n alors T_n converge en probabilité vers $T(F)$ lorsque $n \rightarrow \infty$.

En conséquence si T est faiblement continue dans un domaine \mathcal{B} étoilé en F contenant les Dirac, si en F la courbe d'influence existe et si $T(F) = \theta$ (paramètre à estimer) :

- . T_n converge vers $T((1 - \epsilon)F + \epsilon\delta_x)$ en probabilité pour la loi $((1 - \epsilon)F + \epsilon\delta_x)$
- .. la quantité $\epsilon T'(F, x_0)$ apparaissant dans (2) fournit lorsque " ϵ est assez petit" une valeur approchée du biais asymptotique de l'estimateur T_n : la limite (sous la loi F_ϵ) de ET_n étant $T(F_\epsilon)$ le biais $T(F_\epsilon) - T(F)$ est équivalent à $\epsilon T'(F, x_0)$. Si T' comme fonction de x est borné, le biais asymptotiquement causé par une contamination de F par une masse ϵ est borné.

(b). On peut aussi considérer la continuité, les points de discontinuités et les hauteurs des sauts, les points d'annulation de T' (si $T'(F, x_0) = 0$, $T((1 - \epsilon)F + \epsilon\delta_{x_0}) - T(F) = o(\epsilon)$) (voir HAMPEL [36], p. 386).

(c). On définit en particulier comme on l'a fait dans le paragraphe 1 :

. La sensibilité globale γ^* en F de la fonctionnelle T :

$$\gamma^*(F) = \sup_x |T'(F, x)|.$$

. La sensibilité λ^* au déplacement local d'une masse donnée :

$$\lambda^* = \sup_{x \neq y} \left| \frac{T'(F, x) - T'(F, y)}{x - y} \right|$$

indice utile pour évaluer les effets locaux d'arrondi ou de regroupement (HAMPEL [36] , p. 389).

Pour terminer et avant de fournir un tableau de valeurs numériques, voici deux commentaires :

1. un grand nombre de problèmes concernant la courbe d'influence restent non résolus. Par exemple on ne dispose pas de conditions très générales d'existence de la courbe d'influence : à tel point que ANDREWS [2] considérant sans doute que le nombre de cas particuliers à envisager serait trop important, fournit les expressions des courbes d'influence de plus de 50 estimateurs sans expliciter les conditions d'existence.

2. Quant à l'utilité de la courbe d'influence, HUBER([49] p. 1052) pense que :
 ... "Hampel's influence function is the most important single heuristic tool for constructing robust estimates with specified properties. One will strive for influence functions which are bounded (to limit the influence of any single "bad" observations), which are reasonably continuous in x (to achieve insensitivity against roundoff and grouping effects) and which are reasonably continuous as a function of F (to stabilize the asymptotic variance of the estimate under small changes of F)..." .

4. VALEURS NUMÉRIQUES.

Nous donnons ci-dessous, un tableau de valeurs numériques extrait de HAMPEL ([36], page 392).

Some Numerical Robustness Properties at the Standard Normal Distribution

Estimateur	qr	σ^2	γ^*	δ^*	λ^*	supGEM
M moyenne	-	1.000	∞	0.00	1.00	∞
R n.sc	+	1.000	∞	0.24	1.00	1.48
R H/L	+	1.047	1.77	0.29	1.41	1.29
M Φ (M-type)	+	1.047	1.77	0.50	1.41	1.28
L Φ (L-type)	-	1.047	1.77	0.00	1.41	∞
L mediane	+	1.571	1.25	0.50	∞	1.74
L 5 % Wins	+	1.014	2.13	0.05	∞	1.46
L 5 % trim	+	1.026	1.83	0.05	1.11	1.30
L 10 % trim	+	1.060	1.60	0.10	1.25	1.26
L 6,68 % trim	+	1.037	1.73	0.07	1.15	1.271
M H (1,5)	+	1.037	1.73	0.26	1.15	1.264
M A (1,5)	+	1.037	1.73	0.50	1.15	1.262
M H_1 (1,5)	+	1.037	1.73	0.50	1.15	1.258
M A (1.686)	+	1.024	1.86	0.50	1.10	1.28
M H "sk" 2.71	+	1.066	2.89	0.50	∞	∞
M A (2.71)	+	1.001	2.73	0.50	1.01	1.52

Note : Les colonnes fournissent : qr = robustesse qualitative (et continuité faible qui habituellement coïncide avec elle) ; + = oui, - = non ; σ^2 = variance asymptotique (sous la loi normale ϕ centrée réduite) ; γ^* = sensibilité à la contamination ; δ^* = point de rupture ; λ^* = sensibilité au déplacement local ; supGEM = suprémum de la variance asymptotique dans le modèle de HUBER de (contamination symétrique à 5 %) qui est l'ensemble des $F(x) = 0,95 \phi(x) + 0,05 H(x)$ où H est une distribution symétrique quelconque.

Les lignes fournissent les différents estimateurs : moyenne = moyenne arithmétique ; n.sc = estimateur des scores normaux (R-estimateur déduit du test des scores normaux pour deux échantillons) ; H/L = estimateur de HODGES-LEHMANN (= médiane de toutes les moyennes des paires d'observations = R estimateur déduit du test de WILCOXON) ; ϕ (M-type) = M-estimateur avec $\psi(x) = \phi(x) - 1/2$ (ϕ étant la fonction de répartition de la normale centrée réduite) et le facteur d'échelle est : $(0,6745)^{-1} \times (\text{médiane des écarts absolus à la médiane})$; ϕ (L-type) = L-estimateur équivalent au ϕ (M-type) qui a une courbe d'influence proportionnelle à $\phi(x) - 1/2$ et se comporte très semblablement au L-estimateur optimal pour la distribution logistique ; médiane ; 5 % wins = moyenne 5 % winsorisée ; α % trim = moyenne α % tronquée ; H (k) = proposition (2) de HUBER [44] la valeur du paramètre étant k ; A (k) = "HUBER robust scale" = estimateur de HUBER utilisant comme estimateur du paramètre d'échelle (médiane $|x_i - 50\%|$) / 0,6745 (ANDREWS [2] p. 12) et comme paramètre k ; $H_1(k)$ = estimateurs de HUBER pour une variance connue égale à 1 ; H "sk" k = Moyenne de type "HUBER-skipped" (moyenne avec règle de rejet) définie comme la solution la plus proche de la médiane de $\int \psi\left(\frac{n-\theta}{s}\right) F(dx) = 0$ où $s = (0,6745)^{-1} \times (\text{médiane des écarts absolus à la médiane})$ et $\psi(x) = x$ pour $|x| \leq k$ et $\psi(x) = 0$ pour $|x| > k$.

Les résultats de ce tableau sont basés sur des tables de HAMPEL et des travaux non publiés de HAMPEL ; quelques une des entrées telles que H"sk" 2,71 ont nécessité des travaux particuliers ; la dernière décimale peut-être légèrement inadéquate dans certains cas.

On pourra se reporter au Chapitre IX pour les définitions de δ^{**} et q.r.

Remarque 1 : Pour des compléments concernant les aspects numériques de ces problèmes on pourra se reporter au Chapitre XV .

Remarque 2 : Compte-tenu du nombre de critères : il ne semble pas possible de choisir un estimateur qui soit supérieur à tous les autres. On peut, pour faire un choix, remarquer qu'un estimateur possédant la propriété de robustesse qualitative ($qr = +$) est d'autant meilleur que σ^2 est voisin de 1, γ^* (sensibilité globale), λ^* (sensibilité locale), $\sup GEM$, $\inf GEM$ sont petits et δ^* (point de rupture) grand.

Remarque 3 : Pour HAMPEL [34], le conflit principal concerne σ^2 et γ^* ; on peut vérifier dans le tableau précédent que γ et γ^* varient (presque) en sens contraire.

Claude DENIEAU
Georges OPPENHEIM
Claude VIANO
UER de Mathématique
Université Paris V
12, rue Cujas
75005 PARIS

CHAPITRE XV

ESTIMATEURS D'UN PARAMÈTRE DE POSITION : ASPECTS NUMÉRIQUES

par

Philippe JOLIVALDT - Philippe TRÉCOURT

0 - INTRODUCTION.

A la suite des exposés sur les différents types d'estimateurs d'un paramètre de position proposés pour leur robustesse à des perturbations de nature quelconque, on présente ici quelques résultats relatifs aux trois familles M. L. R.

Les travaux sur la robustesse des estimateurs se sont intéressés essentiellement à déterminer des estimateurs satisfaisant à des critères d'optimalité asymptotique (et minimax). Le comportement de ces estimateurs pour des échantillons de taille finie, qui est essentiel pour le praticien, n'a jusqu'à présent, guère reçu d'éléments de réponse d'où l'intérêt de l'étude faite dans le séminaire de Princeton en 1971 [2] .

De par la nature symétrique des distributions, de par l'équivariance des estimateurs considérés, on est conduit à étudier principalement la variance de ces estimateurs face à des distributions mettant en relief leurs qualités et, (ou) leurs défauts.

Les comparaisons faites dans le cadre de famille d'estimateurs sont extraites, en partie du séminaire de Princeton. Ces résultats ont tous été obtenus par des méthodes de Monté Carlo. Nous avons essayé d'apporter quelques éléments complémentaires pour les L-estimateurs et pour des familles de distributions se prêtant à des

calculs exacts.

Devant la grande variété des estimateurs que l'utilisateur peut rencontrer et afin de préciser des choix parmi les différentes familles d'estimateurs on essaye de dégager des sous-ensembles assez cohérents de distributions dans lesquels un utilisateur accepterait de se placer. L'exposé se présente comme suit :

- Trois attitudes possibles sont décrites,
- les définitions des estimateurs sont rappelées,
- on expose la méthode de calcul,
- les résultats simulés et ceux calculés de façons exactes sont donnés,
- quelques éléments de comparaison sont présentés.

Notations :

$N(0, 1)$	signifie distribution normale centrée réduite
$k N$	signifie distribution normale centrée d'écart type k
$D.EX$	signifie distribution double exponentielle de densité $f(x) = \frac{e^{- x }}{2}$
t_3	signifie distribution de Student à 3 degrés de liberté de densité $f(x) = \frac{1}{\sqrt{\pi} B(\frac{1}{2}, \frac{3}{2})} \frac{1}{(1 + \frac{x^2}{3})^2}$
C	signifie distribution de Cauchy de densité $f(x) = \frac{1}{\pi(1+x^2)}$.

1 - LES ATTITUDES POSSIBLES.

A distance finie, il n'existe pas, à notre connaissance, d'estimateur optimal d'un paramètre de centralité pour toutes les distributions symétriques. On est donc amené à envisager des sous-ensembles de lois dans lesquels la comparaison semble possible.

Utilisant la méthode de comparaison du séminaire de Princeton, on compare les estimateurs face à deux distributions. Le choix des sous-ensembles de lois desquels on extrait ces distributions est appelé une attitude.

Une attitude, c'est donc choisir deux sous-ensembles de lois.

Le premier dit "Modèle de base" représente une famille de lois considérée comme la plus "possible".

Le deuxième dit "perturbé" représente une famille en dehors de laquelle on ne saurait se placer. Ce modèle contient, le plus souvent, le "modèle de base".

Les comparaisons se font pour deux distributions extraites des familles choisies.

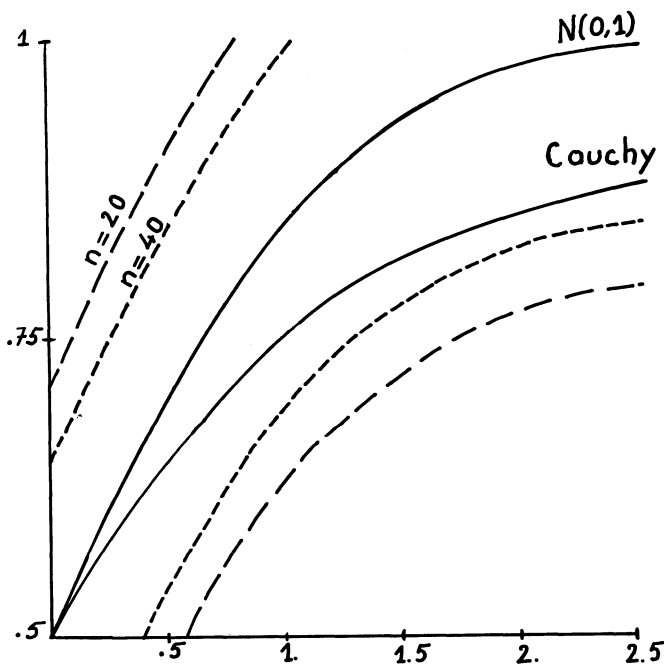
Trois attitudes sont suggérées

	Modèle de base	Modèle "perturbé"
A_1	Loi normale	Modèle de contamination Normale par Normale $((1-\epsilon)N(0,1)+\epsilon N(0,\sigma^2))$ tel que l'écart type de la loi contaminée soit inférieur à deux.
A_2	Loi normale	Même type de voisinage qu'en A_1 mais fort pourcentage de contamination d'écart type plus important.
A_3	étant fixé un pourcentage $\alpha\%$ (modéré) et un estimateur T_0 , on considère un sous-ensemble fini de l'ensemble des distributions β -contaminées (où $\beta \leq \alpha$) de $(0,1)$, la contaminante étant quelconque. On choisit dans ce sous-ensemble la distribution pour laquelle la variance de T_0 est la plus élevée (on l'appelle "la pire des $\alpha\%$ ").	Même type de voisinage qu'en A_1 mais fort pourcentage de contamination d'écart type plus important.

Si l'on cherche à formaliser la notion de voisinage introduite on est amené à considérer des métriques utilisées en robustesse.

Ainsi au voisinage d'une loi normale, on trouve des lois aussi diverses que celle de Cauchy. D'ailleurs en prenant un intervalle de confiance à 70 % même pour une taille d'échantillon $n = 40$ le test de Kolmogoroff ne permet pas de rejeter une distribution type Cauchy au voisinage d'une distribution normale. (Cf. figure 1).

Intervalle de confiance à 70 %



Les traits pleins sont les fonctions de répartition d'une $N(0,1)$ et d'une Cauchy.

Les pointillés sont les bornes d'un intervalle de confiance à 70 % du Test de Kolmogoroff-Smirnov pour une hypothèse de normalité $N(0,1)$ et deux tailles d'échantillon.

Figure 1

De telles métriques ne semblent pas assez discriminantes. On tente alors de caractériser par les choix suivants de distributions les attitudes envisagées.

Attitude A_1 : Dans le modèle de base on prend la distribution $N(0,1)$ et dans le modèle perturbé la double exponentielle.

Attitude A_2 : Dans le modèle de base on prend la distribution $N(0,1)$ et dans le modèle perturbé la distribution de Student à 3 degrés de liberté t_3 .

Attitude A_3 : Dans le modèle de base on prend une loi $N(0,1)$ contaminée par 3 % de distribution quelconque. (En fait les méthodes de calculs de simulation ne prennent en considération pour des résultats comparables, 5 % $3N$ et 5 % $10N$). Le modèle perturbé est constitué de la distribution t_3 .

Enfin on adjoint une étude "Pire de 5 %" contre Cauchy. Certes rares semblent être les situations réelles où une distribution de Cauchy intervient. Cependant dans une vue plus théorique, cela peut permettre de compléter l'information sur le comportement des estimateurs étudiés dans des cas extrêmes (avec des queues très importantes).

2 - LES ESTIMATEURS ÉTUDIÉS.

Tous les estimateurs envisagés sont équivariants sous le groupe Homothétie-translation, c'est-à-dire :

$$T(aX+b) = aT(X)+b.$$

Ils ne sont appliqués qu'à des échantillons dont les distributions sont symétriques. Nous ne faisons ici que de les définir rapidement avec leur notation renvoyant aux exposés pour L, pour M, pour R, pour leurs propriétés.

1) L-estimateurs :

Soit $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ le n -échantillon ordonné. Un L-estimateur est une combinaison linéaire de statistiques d'ordre $T(X) = \sum_{i=1}^n \alpha_i X_{(i)}$ avec $\sum_{i=1}^n \alpha_i = 1$.

Les moyennes tronquées s'écrivent :

$$T_{\alpha}(X) = \frac{1}{n(1-2\alpha)} \left\{ \begin{array}{l} [n(1-\alpha)] \\ \sum_{[n\alpha]+2} X_{(i)} + [1-n\alpha+[n\alpha]](X_{([n\alpha]+1)} + X_{([n(1-\alpha)]+1)}) \end{array} \right\}$$

et sont notées " $\alpha\%$ " en fonction du pourcentage de troncature. La médiane est notée Med.

2) M-estimateur :

Par généralisation de ce que l'on fait pour rechercher les estimateurs par la méthode du maximum de vraisemblance, d'un couple (moyenne, écart-type) pour des lois normales, Huber [44] propose d'étudier un couple (T, S) d'estimateurs solution d'un système du type suivant

$$\begin{cases} \sum_{i=1}^n \psi\left(\frac{X_i - T}{S}\right) = 0 & \text{où } \psi \text{ est une fonction impaire} \\ \sum_{i=1}^n \chi\left(\frac{X_i - T}{S}\right) = 0 & \text{où } \chi \text{ est une fonction paire.} \end{cases}$$

Il faut noter que les propriétés d'un tel couple sont encore très mal connues et que les propriétés asymptotiques ne sont démontrées que dans le cas où la première équation est considérée, S étant remplacé par une estimation.

Deux exemples de tels estimateurs sont étudiés.

a) Type Huber (proposal 2 Huber [44]).

On résoud itérativement le système en partant pour $T_0 = \text{Med}$

$$S_0 = \frac{F_n^{-1}(3/4) - F_n^{-1}(1/4)}{1.35} \quad \text{où } F_n \text{ désigne la fonction de répartition empirique,}$$

$$\text{avec } \psi_k(x) = \begin{cases} -k & x < -k \\ x & -k \leq x \leq +k \\ +k & x > k \end{cases} \quad k \in \mathbb{R}_+^*$$

$$\chi_k(x) = \psi_k^2(x) - \int \psi_k^2(x) \varphi(x) dx \quad \varphi \text{ densité } N(0,1).$$

Ils sont notés H 07 $k = 0,7$

H 12 $k = 1,2$

\vdots

H 20 $k = 2$

b) Type Hampel.

On ne résoud que la première équation en estimant directement $S = \text{Med}(|X_i - \text{Med}|)$

et en prenant pour fonction impaire :

$$\psi_{abc}(x) = \text{Sign}(x) \begin{cases} |x| & |x| < a \\ a & a \leq |x| < b \\ \frac{(c-|x|)a}{c-b} & b \leq |x| < c \\ 0 & |x| \geq c \end{cases}$$

<u>Notation :</u>	a	b	c
HMD	2	2	5,5
25A	2,5	4,5	9,5
21A	2,1	4	8,2
17A	1,7	3,4	8,5
12A	1,2	3,5	8

3) R-estimateurs :

$$\text{H.L.} = \text{Hodges Lehman} = \text{Med} \left\{ \frac{X_i + X_j}{2} \right\}, \quad 1 \leq i \leq j \leq k$$

$$\text{B.H.} = \text{Bickel Hodges} = \text{Med} \left\{ \frac{X_{(i)} + X_{(n+1-i)}}{2} \right\}, \quad i \leq n/2.$$

3 - MÉTHODE DE CALCUL.

Afin de faciliter les calculs numériques, d'améliorer la précision des méthodes de simulation et de profiter du fait que l'on veut comparer tout un ensemble d'estimateurs entre eux, on s'appuie sur deux points de méthode.

a) L'utilisation de variables aléatoires quotients de 2 variables aléatoires indépendantes, ceci revenant alors à travailler sur des lois gaussiennes.

b) L'utilisation de l'équivariance des estimateurs qui permet de diminuer la variabilité des quantités à calculer, i.e. $T(ax+b) = aT(x) + b$.

1) Utilisation des variables aléatoires quotients.

Soit $X = (X_1, X_2, \dots, X_n)$ un n -échantillon de loi F de densité f . Les X_i sont formés comme quotients $X_i = \frac{Z_i}{Y_i}$ où le couple (Z_i, Y_i) est indépendant pour tout i et où Z_i suit une loi $N(0, 1)$.

a) Ainsi, suivant la loi de Y , la loi de $X = Z/Y$ est donné par le tableau suivant :

Loi de Y	Loi de $X = Z/Y$
$Y = \hat{Y} $ où \hat{Y} suit une loi $N(0, 1)$	\rightarrow Loi de Cauchy
Y de densité $f(y) = y^{-3} \exp(-y^2/2)$	\rightarrow Loi D. EX.
Y ponctuelle telle que $\begin{cases} P(Y=1)=1-\alpha \\ P(Y=1/\sigma)=\alpha \end{cases}$	\rightarrow Loi gaussienne contaminée (modèle de Tuckey) par $\alpha\%$ d'une σN

b) Conditionnellement au $(y_i)_{1 \leq i \leq n}$, la loi de $X = (X_1, \dots, X_n)$ est une loi $N(0, v_i^2)$ où $v_i = \frac{1}{y_i}$.

De plus, conditionnellement à $v = (v_1, \dots, v_n)$ les statistiques

$$\hat{X} = \frac{\sum \frac{X_i}{v_i}}{\sum \frac{1}{v_i^2}} \quad \text{qui suit une loi } N\left(0, \frac{1}{\sum \frac{1}{v_i^2}}\right)$$

$$\hat{S}^2 = \frac{\sum \frac{X_i - \hat{X}}{v_i}}{n-1} \quad \text{qui suit la loi d'un } \frac{X_{n-1}^2}{n-1}$$

sont indépendantes et exhaustives.

2) Utilisation de l'équivariance.

Tous les estimateurs étudiés sont équivariants ; on utilise au lieu de l'échantillon aléatoire primitif, la variable aléatoire vectorielle notée $C = \frac{\hat{X} - \hat{X}^*1}{\hat{S}}$ où $1 = (1, 1, \dots, 1)$ qui résume toute l'information et la normalisation de cette variable améliore la précision des résultats simulés.

Pour tout estimateur équivariant T , on a

$$T(X) = (T(\frac{\hat{X} - \hat{X}^*1}{\hat{S}})) \hat{S} + \hat{X} = T(C)\hat{S} + \hat{X} \quad (E).$$

Le calcul de la variance d'un estimateur T et plus généralement de la covariance de deux estimateurs T , U se fera en deux étapes :

- une espérance conditionnellement aux variables V_i et C ,
- une seconde espérance calculée celle-ci par simulation et éliminant le conditionnement puisque :

$$\begin{aligned} \text{Cov}(T(X), U(X)) &= E(\text{Cov}(T(X), U(X) | V, C) + \text{Var}(E(T(X)U(X)) | V, C)) \\ &= E(\text{Cov}(T(X), U(X) | V, C) + 0 \text{ car } T \text{ et } U \text{ centrés}). \end{aligned}$$

On a alors en utilisant la propriété d'équivariance (E)

$$\begin{aligned} \text{Cov}(T(X), U(X) | V, C) &= E\{(T(C)\hat{S} + \hat{X})(U(C)\hat{S} + \hat{X}) | V, C\} \\ &= E\{(T(C)U(C)\hat{S}^2 + (T(C)+U(C))\hat{S}\hat{X} + \hat{X}^2 | V, C\}. \end{aligned}$$

Par linéarité de l'espérance conditionnelle il faut considérer les trois termes suivants :

$$\begin{aligned} E(T(C)U(C)\hat{S}^2 | V, C) &\stackrel{(*)}{=} T(C)U(C)E(\hat{S}^2 | V, C) \stackrel{(**)}{=} T(C)U(C) \\ E((T(C)+U(C))\hat{S}\hat{X} | V, C) &\stackrel{(***)}{=} (T(C)+U(C))E(\hat{S} | V, C)E(\hat{X} | V, C) = 0 \\ E(\hat{X}^2 | V, C) &= \frac{1}{\sum_{i=1}^2 \frac{1}{V_i^2}} \end{aligned}$$

(*), (**) et (***) sont justifiées car conditionnellement à σ les variables \hat{X} , \hat{S}_x et C sont indépendantes.

L'espérance de $\frac{1}{\sum \frac{1}{V_i^2}}$ est en général facile à calculer analytiquement, et avec une très bonne précision à partir des lois des Y_i . La simulation n'intervient donc que pour estimer l'espérance de $T(C)U(C)$.

4 - RÉSULTATS.

Parmi les divers éléments de comparaison on retiendra la convergence vers la normalité largement développée dans ce séminaire, la facilité des calculs précieuse pour le praticien et l'élément essentiel : le graphe de défaut d'efficacité.

1) La convergence vers la normalité a été démontrée pour tous les estimateurs.

Il faudrait en fait mieux connaître le comportement du "Huber proposal 2" qui consiste à travailler sur les deux équations de façon simultanée.

Un "indice de normalité" est défini pour chaque estimateur et chaque probabilité :

$$I. N. (T, F) = \frac{\Psi^{-1}(0,01)}{\varphi^{-1}(0,01)} / \frac{\Psi^{-1}(0,25)}{\varphi^{-1}(0,25)}$$

où Ψ désigne la fonction de distribution de l'estimateur T sous la loi F , et φ celle de la loi normale.

Cet indice est très proche de 1 pour nos estimateurs. Il ne s'en éloigne que pour des distributions à grandes queues et même dans ce cas pour $n \geq 20$ il ne différencie guère les estimateurs.

2) Le fait que les calculs soient plus ou moins aisés importe sans doute à un utilisateur. Si un calcul doit être mené à la main il semble hors de propos de calculer des "Huber" ou des "Hampel". On préférera les moyennes tronquées ou le Bickel-Hodges. Le premier pas d'itération pour le calcul des M-estimateurs semble aussi bien s'effectuer. Si on utilise un calculateur le temps de calcul n'est pas un problème important. Tout au plus peut-on émettre l'hypothèse que dans le cadre de problèmes plus complexes comme la régression, les M-estimateurs prendraient un temps notable.

3) Graphe de défaut d'efficacité.

Les résultats tirés du livre et prêtant à comparaison concernent des échantillons de taille $n = 20$. Pour comparer dans chacune des attitudes types les estimateurs on

peut utiliser non pas la variance mais une grandeur relative et variant dans le même sens que la variance : le défaut d'efficacité d'un estimateur T ,

$$D.E. (T, F) = 1 - \frac{\text{Variance en } F \text{ du "Meilleur" estimateur}}{\text{Var}_F T}.$$

La notion de variance du "Meilleur" estimateur se définit de la manière suivante :

On se donne un ensemble d'estimateurs. Dans le cadre de cette étude il s'agit

- des moyennes α tronquées avec $\alpha = 5\%, 10\%, 15\%, 25\%, 50\%$
- des M-estimateurs type "Huber" $H 07, H 10, H 12, H 15, H 17, H 20$
type "Hampel" $12 A, 17 A, 21 A, 25 A$
- des estimateurs déduits des tests de rang HL et BH.

On appelle pour chaque distribution F étudiée, "Meilleur" estimateur (en F) celui qui dans l'ensemble ci-dessus précisé, a la plus petite variance pour la loi F et c'est cette variance qui est appelée : variance du "Meilleur" estimateur.

Sur les graphiques l'abscisse d'un estimateur est le D.E. dans la situation de base, l'ordonnée le D.E. dans l'alternative.

La précision des résultats liée à la méthode de simulation ne permet pas de distinguer deux points situés dans un carré de côté 1%.

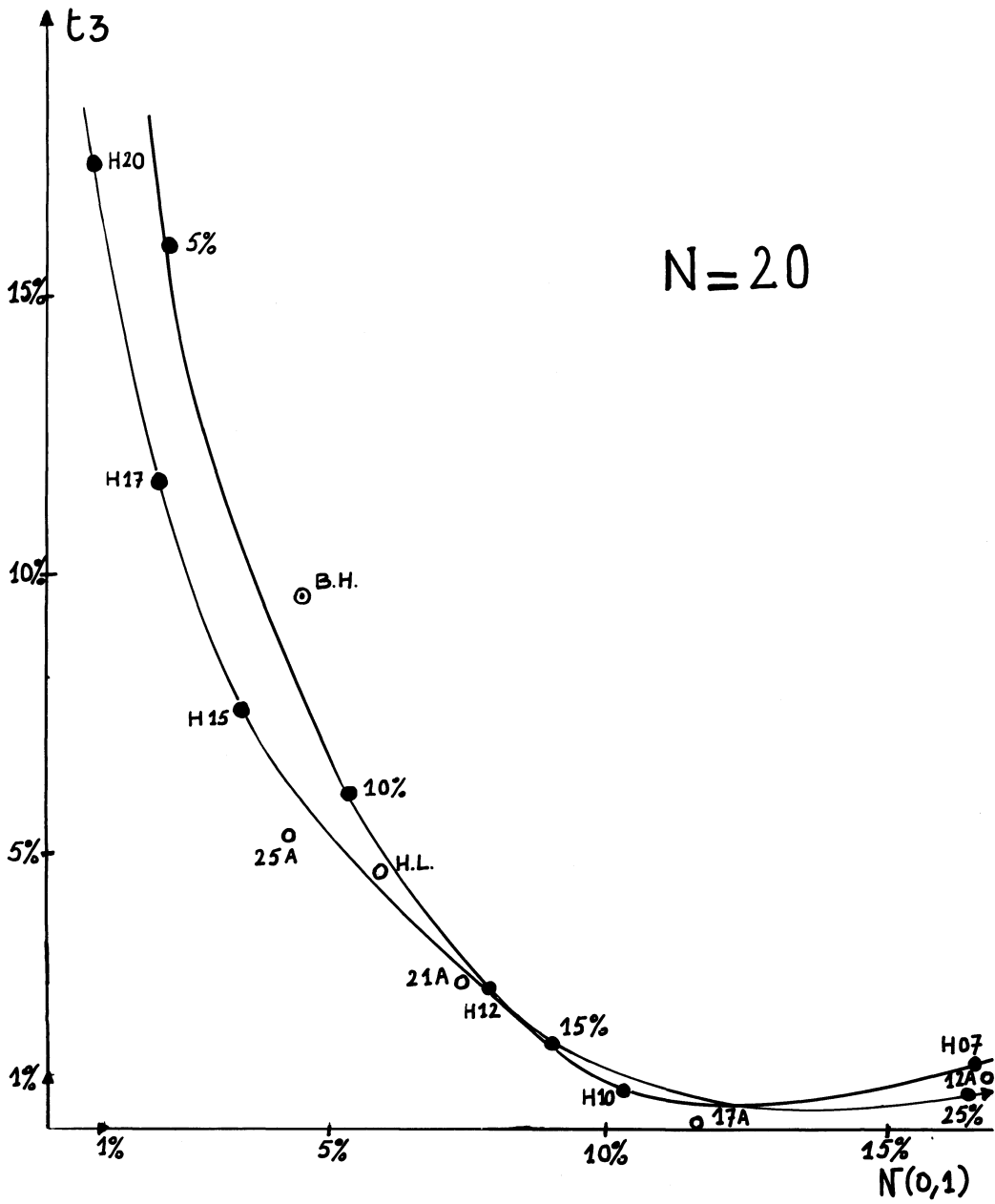
Les graphes des D.E. pour les familles d'estimateurs à 1 paramètre sont continus et même dérivables par morceaux. On peut interpoler entre deux points pour localiser la position d'autres estimateurs, cf. figures 2, 3, 4, 5.

"Où est l'infini ?"

Le tableau suivant donne pour les tailles d'échantillon $n = 10, 20, 40$ ainsi que $n = \infty$, les variances de trois types de troncatures $10\%, 25\%, \text{Med}$ les comparaisons s'établissent pour quatre distributions $N(0, 1), 5\% 2N, 20\% 4N$, D.EX.

(Ce tableau se trouve après les 4 graphiques).

Figure 2



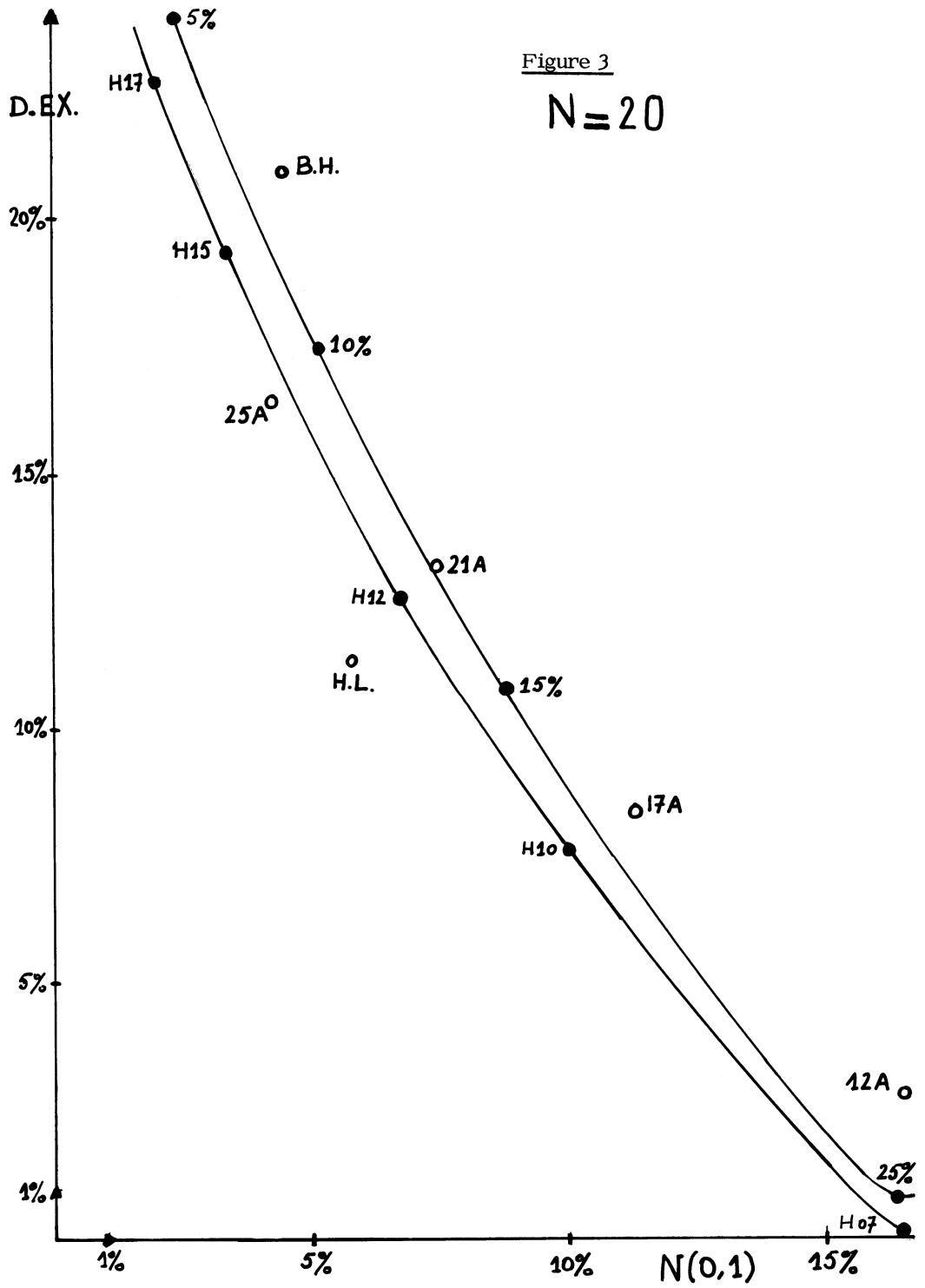
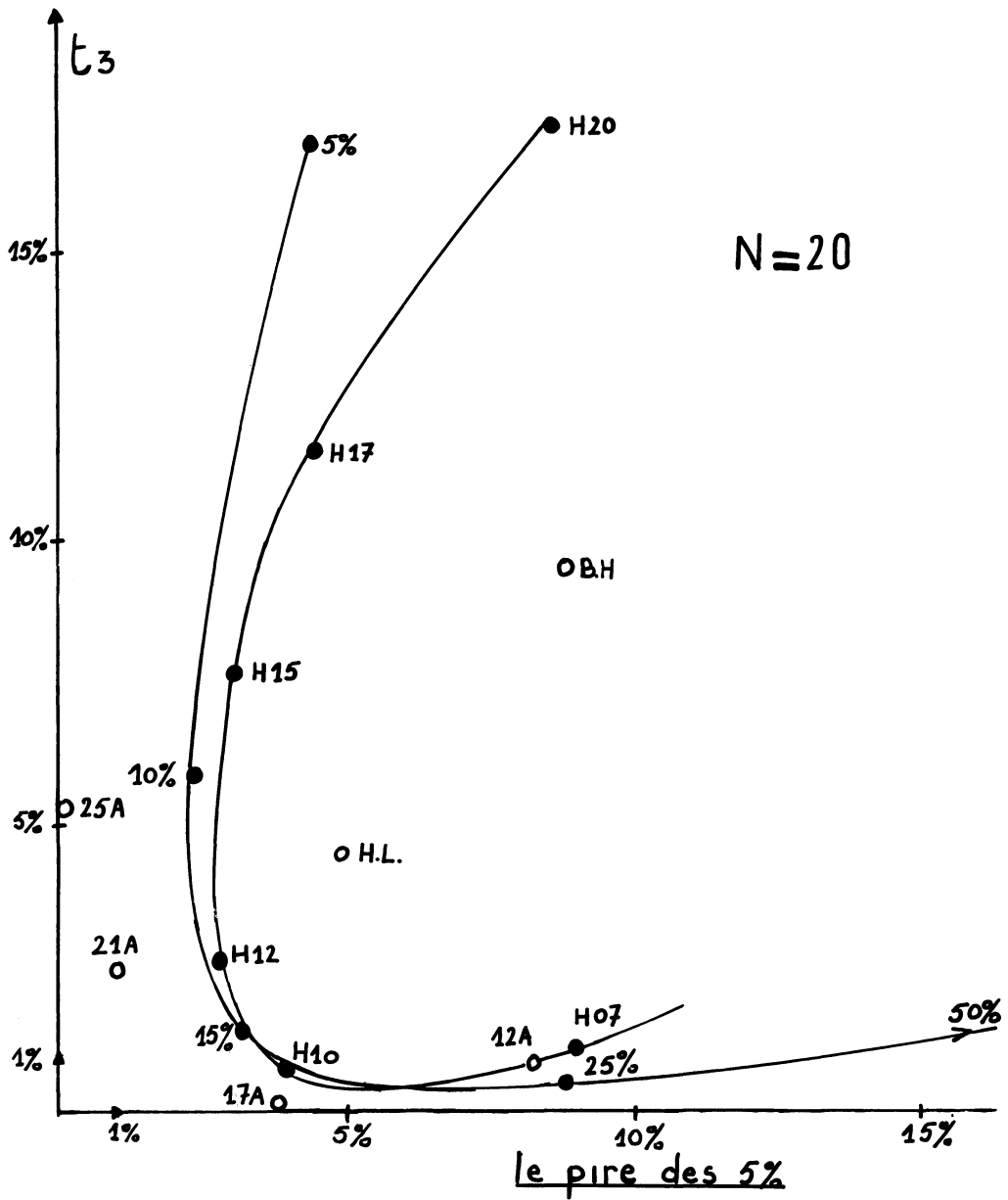
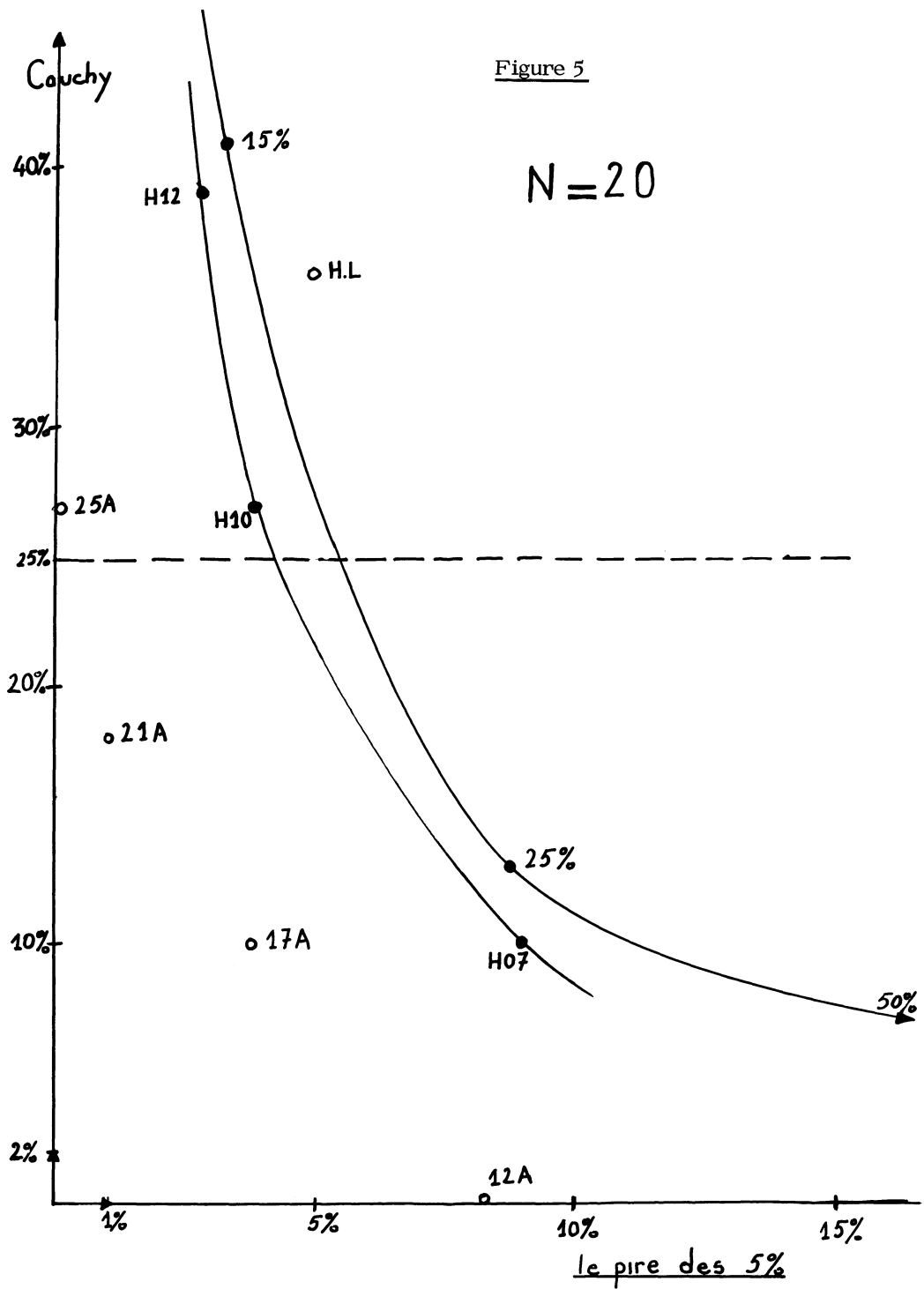


Figure 4





\sqrt{n} Variances des moyennes tronquées

On a calculé et inclus dans ce tableau la meilleure combinaison linéaire des statistiques d'ordre notée BL, et la meilleure moyenne tronquée notée BT (avec entre parenthèses le degré de troncature).

		n = 10	20	40	∞
N(0, 1)	10 %	1,0535	1,0553	1,0573	1,0605
	25 %	1,1643	1,1856	1,1896	1,1952
	Med	1,3833	1,4687	1,5176	1,5708
5 % 2N	10 %	1,1388	1,1343	1,1334	1,1339
	25 %	1,2391	1,2568	1,2590	1,2633
	Med	1,4638	1,5492	1,5983	1,6524
	BL	1,1228	1,1143	1,1089	?
	BT	1,1229 (6 %)	1,1148 (6 %)	1,1093 (4 %)	1,1044 (3 %)
20 % 4N	10 %	2,3885	2,1126	1,9513	1,7960
	25 %	1,9078	1,7927	1,7583	1,7338
	Med	2,0504	2,0905	2,1275	2,1741
	BL	1,8579	1,7218	1,6524	?
	BT	1,8999 (28 %)	1,7888 (21 %)	1,7435 (20 %)	1,7058 (18 %)
D. EX	10 %	1,6166	1,5565	1,5253	1,4941
	25 %	1,4239	1,3278	1,2776	1,2274
	Med	1,4522	1,3324	1,2413	1
	BL	1,3991	1,2740	1,1893	1
	BT	1,4030 (34 %)	1,2766 (37 %)	1,1912 (39 %)	1 (Med)

En fonction de la taille de l'échantillon on peut faire ressortir les points suivants :

- a- Il est certain que pour $n = 10$ on ne peut utiliser les résultats asymptotiques pour les variances ;
- b- Pour $n = 20$ on peut utiliser de tels résultats que si la distribution est très proche de la normale et pour des troncatures inférieures à 25 % ;
- c- Pour les distributions envisagées, en dehors de la médiane les valeurs asymptotiques

tiques sont quasiment atteintes pour $n = 40$.

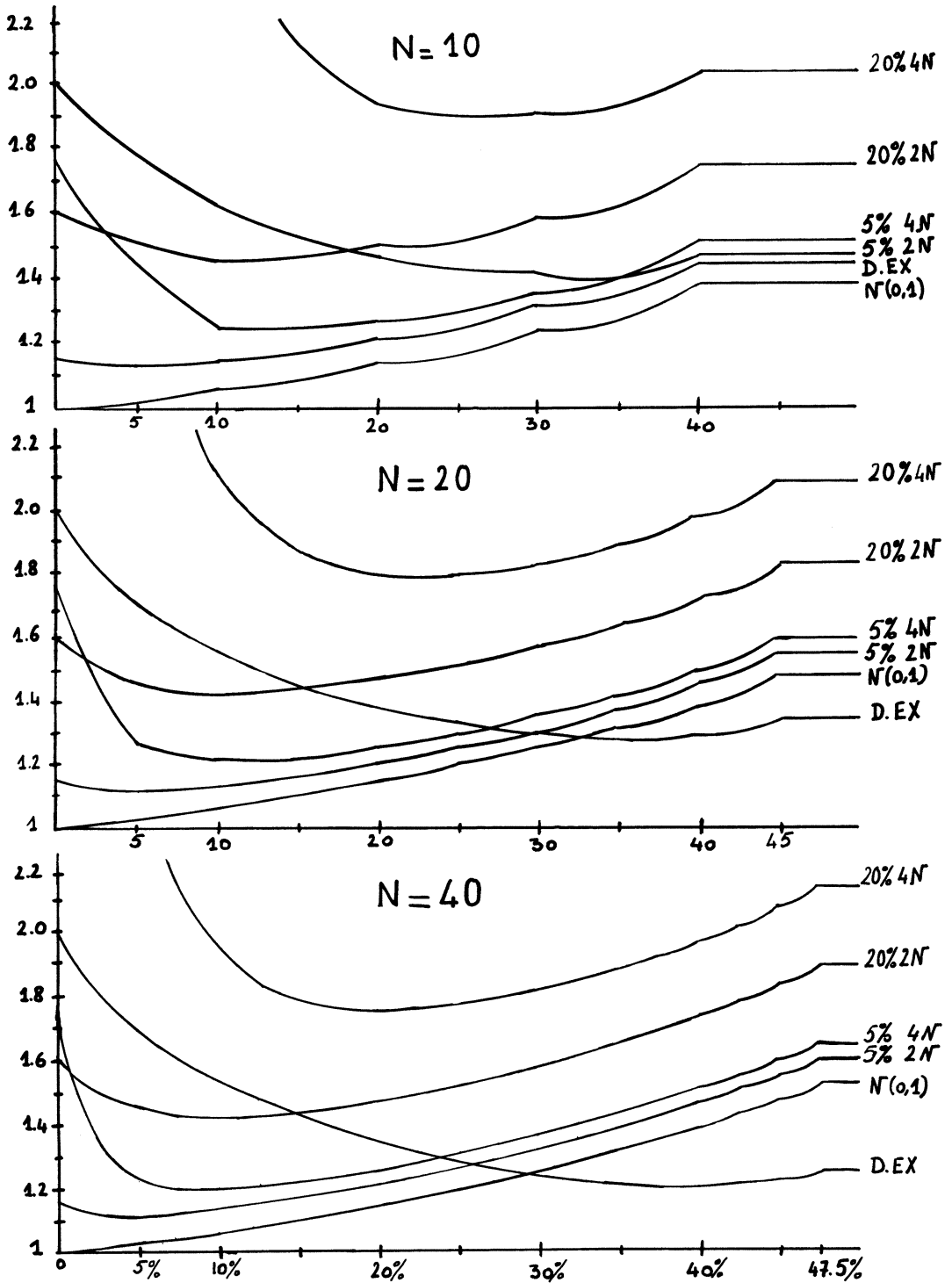
Il faut remarquer que pour la D. EX, la médiane asymptotiquement optimale est encore loin de la valeur à l'infini pour $n = 40$. De plus une moyenne tronquée à 35 %, pour $n = 40$, possède une variance de 1,2014 donc plus faible.

Nous devons cependant ajouter en ce qui concerne la médiane qu'en réalité on ne calcule jamais, sauf dans le cas asymptotique, une médiane mais toujours une moyenne tronquée de pourcentage proche de 50 %. Ainsi pour $n = 40$,

$$\text{Med} = \frac{X_{(20)} - X_{(21)}}{2} = 47,5 \text{ \%} .$$

la variance d'une moyenne tronquée sera proche de sa valeur asymptotique dès que l'on conservera plus d'une dizaine de données (au moins dans les limites de cette étude).

La figure 6 présente pour trois tailles d'échantillons $n = 10, 20, 40$ et pour 6 distributions les variances multipliées par \sqrt{n} des moyennes tronquées à α % .

Figure 6 - \sqrt{n} variance des α % moyennes tronquées.

5 - CONCLUSION.

A - Quelques éléments de conclusion du séminaire de Princeton.

Les participants sont convaincus de l'intérêt d'étudier des familles d'estimateurs à (au moins) un paramètre réel et

a) de s'efforcer de choisir "au mieux" la valeur du paramètre dans chaque famille,

b) de comparer ces familles entre elles.

De plus ces comparaisons sont plus efficaces si elles se font dans le cadre de deux situations simultanées ce qui conduit au graphe de défaut d'efficacité.

Persuadés qu'aucun estimateur n'est valable quelle que soit la distribution de départ (ce qui est vérifié pour les tailles considérées) ils ont cherché à classer ces estimateurs en fonction de différentes attitudes envisageables.

Il leur semble possible d'utiliser les résultats asymptotiques dès que la taille de l'échantillon dépasse 20 (sauf pour des distributions à grosses queues type Cauchy).

Par ailleurs, les estimateurs les plus intéressants semblent être à leur avis :

les "Hampel"

les "Huber".

Une promesse "du côté des adaptatifs" ; les "Hampel" et les "Huber" paraissent d'ailleurs intéressants dès que l'on s'en tient au 1^{er} pas de l'itération nécessaire à leur calcul.

B - Quelques remarques complémentaires que nous apportons.

Les calculs sur les L-estimateurs n'ont pu être fait que dans le modèle de contamination normale-normale et quelques lois pour lesquelles on pouvait faire les calculs. Ce modèle de contamination ne cherche absolument pas à recouvrir toute la réalité mais néanmoins permet de mieux comprendre certains problèmes liés

à la robustesse.

Le fait de comparer des estimateurs dans le cadre de famille et face à deux distributions simultanées semble fructueux. Il nous est cependant apparu nécessaire lors de l'étude un peu plus détaillée des moyennes tronquées de préciser les limites d'un problème de robustesse. Les qualités d'un estimateur sont intéressantes si l'on se place à proximité d'une distribution.

Qu'entend-on par proximité ?

1) La contamination n'est pas trop importante. Considérer des distributions où la contamination est de 50 % peut certes être rendu obligatoire par le problème posé mais ne relève plus de la robustesse.

Dans un tel cas il faut développer une connaissance plus approfondie des données recueillies et des méthodes susceptibles de les traiter.

2) La distribution contaminante n'a pas de queues trop importantes. Sinon une étude des données aberrantes permet de ramener l'échantillon dans des bornes plus raisonnables.

Il nous semble préférable de chercher les données aberrantes et leurs causes plutôt que d'intégrer dans l'estimateur une procédure de rejet.

- Les résultats asymptotiques auront d'autant plus de valeurs que des études précises sur la vitesse de convergence permettront de comparer les estimateurs sur des familles de lois les plus larges possibles.

Face à ces positions, on considère dans les calculs faits des contaminations d'au plus 20 % par des distributions d'écart type modéré < 4 . Parmi les estimateurs étudiés les moyennes tronquées sont préférées pour les raisons suivantes :

- dans l'ensemble des L-estimateurs, elles sont proches de l'optimum et des troncatures entre 15 et 25 % sont assez stables pour les distributions envisagées. Dans l'ensemble M, L, R elles sont les plus aisées et les plus rapides à calculer.

Philippe JOLIVALDT
Département de Statistique
Université Paris I
Place du Panthéon
75231 PARIS

Philippe TRÉCOURT
Chaire de Mathématiques et
Informatique
Institut National Agronomique
Paris-Grignon
16, rue Claude Bernard
75005 PARIS

CHAPITRE XVI

ESTIMATEURS ADAPTATIFS DANS UN MODÈLE DE REGRESSION

par
 Dominique PICARD

1 - EXPOSÉ DU PROBLÈME.

On considère le modèle de régression :

$$Y_{ij} + \theta_i + \alpha X_{ij} = E_{ij} \quad \begin{array}{l} j \text{ variant de } 1 \text{ à } n_i \\ i \text{ variant de } 1 \text{ à } k . \end{array}$$

Les E_{ij} sont les variables observées (indépendantes 2 à 2) . Les Y_{ij} sont des variables aléatoires de lois F_i inconnues, supposées symétriques autour de l'origine.

$\Theta = (\alpha, \theta_1, \dots, \theta_k)$ est le vecteur des paramètres inconnus que l'on cherche à estimer. Les X_{ij} sont des facteurs constants.

L'objet du problème est de trouver un estimateur Θ_n de Θ non randomisé, équi-variant par translation et homothétie , asymptotiquement efficace.

Nous étudierons les cas où les lois F_i admettent une densité f_i par rapport à la mesure de LEBESGUE, et d'information de FISHER finie.

On est en situation non paramétrique.

On va adopter un point de vue adaptatif c'est-à-dire se servir de l'échantillon

pour estimer les distributions F_i (après une estimation grossière de θ) puis dans le modèle estimé, on traitera le problème de l'estimation de θ , de façon paramétrique.

Nous allons reprendre les méthodes de STONE exposées dans [87] sur l'estimation d'un paramètre de position et les étendre au modèle de régression.

2 - EXPOSÉ DE LA MÉTHODE.

On va d'abord estimer Θ à l'aide d'un estimateur $\bar{\Theta}_n = (\bar{\alpha}, \bar{\theta}_1, \dots, \bar{\theta}_k)$ équivariant par translation et homothétie, asymptotiquement normal (i.e.

$n^{1/2} (\bar{\Theta}_n - \Theta) \xrightarrow{L} \mathcal{N}(0, \sigma^2)$ quand $n \rightarrow +\infty$).

(Des exemples de tels estimateurs peuvent être trouvés dans l'article de YOHAI [100]).

On peut ensuite estimer F_i (en supposant que $E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}$ suit approximativement la loi F_i).

L'estimation la plus courante de F_i est \hat{F}_{n_i} la répartition empirique :

$$\hat{F}_{n_i}(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} 1_{\{(E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}) < x\}}$$

Dans la mesure où l'on est intéressé par des propriétés de régularité de f_i , on est amené à prendre comme estimateur de f_i le convolué de \hat{F}_{n_i} avec un noyau régularisant (par exemple un noyau gaussien noté $\varphi(x, r) = \frac{1}{(2\pi)^{1/2} r} \exp - \frac{x^2}{2r^2}$). Enfin pour garder les propriétés dues à la symétrie de f_i , on prend le symétrisé de l'estimateur obtenu :

$$\begin{aligned} \hat{f}_{n_i}(x, F_{n_i}) &= \frac{1}{2} \left[\frac{1}{n_i} \sum_{j=1}^{n_i} \varphi(x - E_{ij} + \bar{\theta}_i + \bar{\alpha} X_{ij}; r_{n_i}) \right] \\ &+ \frac{1}{n_i} \sum_{j=1}^{n_i} \varphi(-x - E_{ij} + \bar{\theta}_i + \bar{\alpha} X_{ij}; r_{n_i}) \end{aligned}$$

Nous allons ensuite construire, dans un modèle paramétrique (c'est-à-dire les F_i étant supposées connues), un estimateur $\tilde{\Theta}_n$ équivariant par translation et homothétie, non randomisé, asymptotiquement efficace (ceci quelles que soient les distributions F_i de départ).

Enfin on s'assure que le modèle estimé est, du moins pour le problème qui nous intéresse, suffisamment voisin du modèle réel, c'est-à-dire que les estimateurs $\tilde{\Theta}$ (obtenus de façon théorique en supposant les F_i connues) et $\hat{\Theta}$ (obtenu en remplaçant dans $\tilde{\Theta}$ F_i par son estimation) sont asymptotiquement proches.

3 - PARTIE NON ADAPTATIVE : CONSTRUCTION D'UN ESTIMATEUR.

Supposons les densités f_i connues. L'estimateur du maximum de vraisemblance est défini comme racine du système d'équations :

$$\begin{cases} \sum_{j=1}^{n_i} L_i (E_{ij} - \theta_i - \alpha X_{ij}) = 0 \\ \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij} L_i [E_{ij} - \theta_i - \alpha X_{ij}] = 0 \end{cases} \quad \text{pour } i = 1, \dots, k$$

(où $L_i(x) = \frac{f_i'(x)}{f_i(x)}$) peut être approché par $\underline{\theta} = (\underline{\alpha}, \underline{\theta}_1, \dots, \underline{\theta}_k)$ où

$$\begin{aligned} \underline{\theta}_i &= \bar{\theta}_i + \frac{\sum_{j=1}^{n_i} L_{ij}}{n_i} - (\underline{\alpha} - \bar{\alpha}) \frac{\sum_{j=1}^{n_i} X_{ij} L'_{ij}}{n_i} \\ &\quad - \frac{\sum_{j=1}^{n_i} L'_{ij}}{\sum_{j=1}^{n_i} L'_{ij}} \\ \underline{\alpha} &= \bar{\alpha} + \frac{\sum_{i=1}^k \left[\sum_{j=1}^{n_i} X_{ij} L_{ij} - \left(\sum_{j=1}^{n_i} X_{ij} L'_{ij} \right) \left(\sum_{j=1}^{n_i} L_{ij} \frac{1}{n_i} \right) \right]}{\sum_{j=1}^{n_i} L'_{ij}} \\ &\quad - \frac{\sum_{i=1}^k \left[\sum_{j=1}^{n_i} X_{ij}^2 L'_{ij} - \frac{\left(\sum_{j=1}^{n_i} X_{ij} L'_{ij} \right)^2}{n_i} \right]}{\sum_{j=1}^{n_i} L'_{ij}} \end{aligned}$$

où L_{ij} et L'_{ij} sont des abréviations respectivement de $L_i(E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij})$ et $L'_i(E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij})$.

On est amené à imposer les conditions suivantes aux facteurs contrôlés X_{ij} :

- * $\sum_{j=1}^{n_i} X_{ij} = 0 \quad \forall i$ (travailler sur des facteurs centrés)
- * $\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}^2 \rightarrow b_i < +\infty$ quand $n_i \rightarrow \infty$, $\forall i \in \{1, \dots, k\}$
- * $\frac{1}{n_i} \sum_{j=1}^{n_i} |X_{ij}|^3 \rightarrow \alpha_i < +\infty$ quand $n_i \rightarrow \infty$, $\forall i \in \{1, \dots, k\}$
- * $\frac{n_i}{n} \rightarrow 1$ quand $n \rightarrow \infty$ pour tout i

$$* \sup_{i,j} |X_{ij}| < M n^{\frac{1-\delta}{4}} \quad \text{avec } 0 < \delta < 1$$

(Ces conditions sont en particulier vérifiées, lorsque le domaine de variation des X_{ij} est borné, ce qui est un cas fréquent).

On a alors :

$$\begin{aligned} \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}^2 L'_{ij} &\xrightarrow{ps} -b_i I(f_i) \\ \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} L'_{ij} &\xrightarrow{ps} 0 \\ \frac{1}{n_i} \sum_{j=1}^{n_i} L'_{ij} &\xrightarrow{ps} -I(f_i) \quad (E L'_{ij}(x) = -I(f_i)) \end{aligned}$$

En remplaçant dans le précédent estimateur ces estimateurs par leurs limites on obtient

$$\begin{aligned} \underline{\alpha} &= \bar{\alpha} - \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij} L_i(E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij})}{\sum_{i=1}^k n_i b_i I(f_i)} \\ \theta_j &= \bar{\theta}_j - \frac{\sum_{i=1}^{n_j} L_i(E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij})}{n_i I(f_i)} \end{aligned}$$

Pour régulariser f_i remplaçons là par $F_i * \varphi(., r_n)$ où φ est le noyau régularisant défini plus haut. On obtient ainsi un nouvel estimateur en remplaçant $L_i(x)$ par $L_i(., r_n) = \frac{f_i * \varphi(., r_n)}{f_i * \varphi(., r_n)}$, f_i par $f_i(., r_n) = f_i * \varphi(., r_n)$.

Cette convolution revient à remplacer l'échantillon E_{ij} par l'échantillon $E_{ij} + W_{ij}$ où les W_{ij} sont des variables aléatoires indépendantes des E_{ij} et deux à deux suivant des lois normales de moyennes 0, de variance r_n^2 (on fera tendre r_n vers 0).

De plus, on tronque ces densités au moyen d'une fonction g à support $(-1, +1)$, de classe C^2 , comprise entre 0 et 1.

Ainsi, on peut approcher $I(f_i)$ par

$$A_i(r_n, c_n) = \int L_i^2(x, r_n) g\left(\frac{x}{c_n}\right) f_i(x, r_n) dx$$

(la densité est tronquée sur $[-c_n, +c_n]$; on fera évidemment tendre c_n vers l'infini).

$\underline{\Theta}$ est remplacé par :

$$\underline{\alpha} = \bar{\alpha} - \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij} L_i(E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}, r_n) g\left(\frac{E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}}{c_n}\right)}{\sum_{i=1}^k \sum_{j=1}^{n_i} b_i A_i(r_n, c_n)}$$

$$\underline{\theta}_i = \bar{\theta}_i - \frac{\sum_{j=1}^{n_i} L_i(E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}, r_n) g\left(\frac{E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}}{c_n}\right)}{n_i A_i(r_n, c_n)}$$

On centre approximativement cet estimateur en lui ajoutant les constantes de centres obtenues au point $\bar{\Theta} = 0$:

$$C_{\alpha} = 0$$

$$C_{\theta_i} = - \frac{1}{A_i(r_n, c_n)} \int L_i(r_n, c_n) g\left(\frac{x}{c_n}\right) f_i(x, r_n) dx$$

(On remarquera que $C_{\theta_i} = 0$ si g est choisie symétrique autour de l'origine).

Enfin, pour obtenir un estimateur non randomisé (éliminer les W_{ij}) et diminuer la variance, on en prend l'espérance conditionnelle par rapport aux E_{ij} soit :

$$\left(\begin{array}{l} \tilde{\alpha} = \bar{\alpha} - \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij} E^{(E_{ij})} L_i(E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}; r_n) g\left(\frac{E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}}{c_n}\right)}{\sum_{i=1}^k \sum_{j=1}^{n_i} b_i A_i(r_n, c_n)} \\ \tilde{\theta}_i = \bar{\theta}_i - \frac{\sum_{j=1}^{n_i} E^{(E_{ij})} L_i(E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}, r_n) g\left(\frac{E_{ij} - \bar{\theta}_i - \bar{\alpha} X_{ij}}{c_n}\right)}{n_i A_i(r_n, c_n)} + C_{\theta_i} \end{array} \right)$$

D'où la forme finale de l'estimateur non adaptatif

$$\tilde{\Theta} = (\tilde{\alpha}, \tilde{\theta}_1, \dots, \tilde{\theta}_n)$$

On notera $\varphi(x)$ pour $\varphi(x, r_n)$ quand il n'y a pas d'ambiguïté.

$$\left\{ \begin{array}{l} \tilde{\alpha} = \bar{\alpha} - \frac{\sum_{i=1}^k n_i \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} \varphi(y - E_{ij} + \bar{\theta}_i + \bar{\alpha} X_{ij}) dy}{\sum_{i=1}^k n_i b_i A_i(r_n, c_n)} \\ \tilde{\theta}_i = \bar{\theta}_i - \frac{\int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \left[\sum_{j=1}^{n_i} \frac{1}{n_i} \varphi(y - E_{ij} + \bar{\theta}_i + \bar{\alpha} X_{ij}) - f(y, r_n) \right] dy}{A_i(r_n, c_n)} \end{array} \right.$$

Considérons les propriétés de cet estimateur :

1) Equivariance :

L'estimateur $\tilde{\theta}$ est évidemment équivariant par translation ; pour obtenir l'équivalence par changement d'échelle, prenons pour (r_n) et (c_n) les suites $(t_n s_n)$ et $(d_n s_n)$ où (s_n) est suite de variables aléatoires, ne dépendant que de l'échantillon E_{ij} , équivariantes par changement d'échelle et tendant en probabilité vers une valeur : $s(n)^{1/2} \left(\frac{s_n}{s} - 1 \right) = O_p(1)$ quand $n \rightarrow \infty$; t_n et d_n sont des suites numériques tendant respectivement vers 0 et l'infini.

Des inégalités du type :

$$\frac{s}{s_n} f(x, t_n s) \leq f(x, t_n s_n) \leq \left(\frac{s}{s_n} \right)^{s^2/s_n^2} f(x, t_n s)$$

et (3), annexe 1) permettent d'éliminer le problème de l'aléat.

2) Efficacité asymptotique :

Sous la condition : il existe $\epsilon > 1 - \delta$ tel que $\frac{c_n}{n^{1-\epsilon} r_n} = O_p(1)$, alors

$$n^{1/2} (\tilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, I^{-1})$$

où

$$I = \begin{pmatrix} \sum_{i=1}^k b_i I(f_i) & 0 & 0 & 0 \\ 0 & I(f_1) & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & I(f_k) \end{pmatrix}$$

Démonstration :

Proposition 1 : $\lim_{\substack{r_n \rightarrow 0 \\ c_n \rightarrow +\infty}} A_i(r_n, c_n) = I(f_i) .$

Ce résultat est démontré par STONE dans [87] .

Proposition 2 :

$$\tilde{\theta}_i - \theta_i = \frac{1}{A_i(r_n, c_n)} \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) [f_{n_i}(y, r_n) - f_i(y, r_n)] dy + o_p(1) n^{-1/2}$$

$$\tilde{\alpha} - \alpha = \frac{\sum_{i=1}^k n_i \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \left[\sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} \varphi(y - y_{ij}, r_n) \right] dy}{\sum_{i=1}^k n_i b_i A_i(r_n, c_n)} + o_p(1) n^{-1/2}$$

$$\text{où } f_{n_i}(y, r_n) = \frac{1}{n_i} \sum_{j=1}^{n_i} \varphi(y - y_{ij}, r_n) .$$

Démonstration : On a :

$$\begin{aligned} \tilde{\theta}_i &= \bar{\theta}_i - \frac{1}{A_i(r_n, c_n)} \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \left\{ (f_{n_i}(y, r_n) - f_i(y, r_n)) \right. \\ &\quad + (\bar{\alpha} - \alpha) \sum_{j=1}^{n_i} \frac{1}{n_i} \varphi'(y - y_{ij}, r_n) + (\bar{\theta}_i - \theta_i) (f'_{n_i}(y, r_n) - f'_i(y, r_n)) \\ &\quad + (\bar{\theta}_i - \theta_i) (f'_i(y, r_n)) + \frac{1}{2} (\bar{\theta}_i - \theta_i)^2 \sum_{j=1}^{n_i} \frac{1}{n_i} \varphi''(y - y_{ij} + t_1(\bar{\theta}_i - \theta_i) + t_2(\bar{\alpha} - \alpha) X_{ij}) \\ &\quad + \frac{1}{2} (\bar{\alpha} - \alpha)^2 \sum_{j=1}^{n_i} \frac{1}{n_i} X_{ij}^2 \varphi''(y - y_{ij} + t_1(\bar{\theta}_i - \theta_i) + t_2(\bar{\alpha} - \alpha) X_{ij}) \\ &\quad \left. + (\bar{\alpha} - \alpha) (\bar{\theta}_i - \theta_i) \sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} \varphi''(y - y_{ij} + t_1(\bar{\theta}_i - \theta_i) + t_2(\bar{\alpha} - \alpha) X_{ij}) \right\} dy \end{aligned}$$

$$\text{où } 0 < t_1 < 1 \quad \text{et} \quad 0 < t_2 < 1$$

Les 2 premiers termes se traitent comme dans l'article de STONE [87] .

$$\text{i.e. : } \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) (f'_{n_i}(y, r_n) - f'_i(y, r_n)) dy = o_p(1)$$

$$\frac{1}{A_i(r_n, c_n)} \int L(y, r_n) g\left(\frac{y}{c_n}\right) f'_i(y, r_n) dy = 1.$$

Il est clair, vues les conditions imposées aux X_{ij} que

$$\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \varphi'(y - y_{ij}) = o_p(1).$$

La majoration des termes du second ordre utilise l'étude faite en annexe sur les comportements des fonctions $\psi_{n_i}(y, r_n, c_{ij})$ et $\psi_{n_i}(y, r_n)$.

On a à étudier des termes de la forme

$$\begin{aligned} \Gamma_i(r_n, c_n) = \frac{1}{A_i(r_n, c_n)} \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} \varphi''(y - y_{ij} + t_1(\bar{\theta}_i - \theta_i) \\ + t_2(\bar{\alpha} - \alpha) X_{ij}) \end{aligned}$$

Posons $c_{ij} = t_1(\bar{\theta}_i - \theta_i) + t_2(\bar{\alpha} - \alpha) X_{ij}$

$$\psi_{n_i}(y, r_n, c_{ij}) = \frac{1}{n_i} \sum_{j=1}^{n_i} |\gamma_{ij}| \varphi(y - y_{ij} + c_{ij}, r_n)$$

D'après (1) annexe 1 et les propriétés des $f_i(x, r_n)$ étudiées par STONE dans [87]

$$\begin{aligned} |\Gamma_i(r_n, c_n)| \leq O_p(1) \int_{-c_n}^{+c_n} \frac{1}{r_n^3} [\psi_{n_i}(y, r_n, c_{ij}) r_n^{-\epsilon} + \psi_{n_i}^{1-\epsilon}(y, r_n, c_{ij})] \times \\ \left[\frac{1}{r_n^\epsilon} + \frac{1}{f_i(y, r_n)^\epsilon} \right] dy \end{aligned}$$

D'après (2), annexe 1, puisque $(\bar{\theta}_i - \theta_i) = O_p(1) n^{-1/2}$ et $(\bar{\alpha} - \alpha) = O_p(1) n^{-1/2}$,

$\sup_{i,j} |X_{ij}| \leq M_1 n^{1/4}$ et $\sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} < M_2$. On a

$$\begin{aligned} |\Gamma_i(r_n, c_n)| \leq O_p(1) \int_{-c_n}^{+c_n} \frac{1}{r_n^3} \left\{ \frac{\psi_{n_i}(y, r_n) + \psi_{n_i}^{1-\epsilon}(y, r_n)}{r_n^\epsilon} + \right. \\ \left. + (\psi_{n_i}(y, r_n) + \psi_{n_i}^{1-\epsilon}(y, r_n))^{1-\epsilon} \right\} \times \left(\frac{1}{r_n^\epsilon} + \frac{1}{f_i(y, r_n)^\epsilon} \right) dy \end{aligned}$$

où $\psi_{n_i}(y, r_n) = \frac{1}{n_i} \sum_{j=1}^{n_i} \gamma_{ij} \varphi(y - y_{ij}, r_n)$

De (7) annexe 2) comme $\frac{1}{n_i^2} \sum_{j=1}^{n_i} \gamma_{ij}^2 < \frac{M}{n^\delta}$ on déduit que s'il existe

$$\epsilon > 1 - \delta / \frac{c_n}{n^{1-\epsilon} r_n^6} = O_p(1).$$

$$\text{Alors} \quad \Gamma_i(r_n, c_n) = O_p(1) \frac{n^\epsilon}{r_n^3} = o_p(1) n^{1/2}.$$

D'où la majoration des termes du second ordre, en prenant successivement $\gamma_{ij} = 1$, $\gamma_{ij} = |X_{ij}|$, $\gamma_{ij} = X_{ij}^2$.

Pour α , on fait une démonstration analogue :

$$\begin{aligned} \tilde{\alpha} - \bar{\alpha} &= \frac{1}{\sum_{i=1}^k n_i b_i A_i(r_n, c_n)} \times \left[\sum_{i=1}^k n_i \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \right. \\ &\quad \left\{ \sum_{j=1}^{n_i} \frac{1}{n_i} X_{ij} \varphi(y - y_{ij}, r_n) + (\bar{\alpha} - \alpha) \left(\sum_{j=1}^{n_i} \frac{1}{n_i} X_{ij}^2 \varphi'(y - y_{ij}, r_n) - \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}^2 f'_i(y, r_n) \right) \right. \\ &\quad \left. - (\bar{\alpha} - \alpha) \sum_{j=1}^{n_i} \frac{X_{ij}^2}{n_i} f'_i(y, r_n) - (\bar{\theta}_i - \theta_i) \left(\sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} \varphi'(y - y_{ij}; r_n) \right) \right. \\ &\quad \left. + \frac{1}{2} (\theta_i - \bar{\theta}_i)^2 \sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} \varphi''(y - y_{ij} + t_{1i}(\bar{\theta}_i - \theta_i) + t_{2i}(\bar{\alpha} - \alpha) X_{ij}; r_n) \right. \\ &\quad \left. + \frac{1}{2} (\alpha - \bar{\alpha})^2 \sum_{j=1}^{n_i} \frac{X_{ij}^3}{n_i} \varphi''(y - y_{ij} + t_{1i}(\bar{\theta}_i - \theta_i) + t_{2i}(\bar{\alpha} - \alpha) X_{ij}; r_n) \right. \\ &\quad \left. + (\bar{\alpha} - \alpha) (\bar{\theta}_i - \theta_i) \sum_{j=1}^{n_i} \frac{X_{ij}^2}{n_i} \varphi''(y - y_{ij} + t_{1i}(\bar{\theta}_i - \theta_i) + t_{2i}(\bar{\alpha} - \alpha) X_{ij}; r_n) \right\} dy \Big] \end{aligned}$$

$$\text{or} \quad \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}^2 f'_i(y, r_n) dy \rightarrow b_i A_i(r_n, c_n)$$

$$\int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \sum_{j=1}^{n_i} \frac{1}{n_i} X_{ij}^2 (\varphi'(y - y_{ij}, r_n) - f'_i(y, r_n)) dy$$

$$\leq \left[\int_{-c_n}^{+c_n} \left(\frac{f_i + f_i^{1+\epsilon}}{f_i^\beta} \right)^2 \right]^{1/2} \left[\int_{-c_n}^{+c_n} \frac{\sum_{j=1}^{n_i} \frac{X_{ij}^2}{n_i} (\varphi'(y - y_{ij}, r_n) - f'_i(y, r_n))^2}{f_i^{2-2\beta}} \right]^{1/2}$$

$$\leq O_p(1) \frac{n^{2\epsilon}}{n^\delta r_n^3} \leq O_p(1) \quad \text{si} \quad \frac{c_n}{n^\delta r_n} = O_p(1) \quad (\text{d'après (9), annexe 2})$$

$\sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} \varphi'(y - y_{ij}, r_n)$ est de façon claire un $o_p(1)$. Les termes du second ordre font appel aux mêmes arguments que le paragraphe précédent.

Proposition 3 : $n^{1/2} (\tilde{\Theta} - \Theta) \xrightarrow{\mathcal{L}} N(0, I^{-1})$ sous la condition

$$\frac{c_n}{n^{1-\epsilon} r_n} = O_p(1) \quad 1 - \delta < \epsilon < 1$$

Il est clair, vu la forme de $\tilde{\Theta}$ dans la proposition 2, que la loi limite est normale de moyenne $(0, \dots, 0)$.

Calculons la matrice de covariance limite :

$$\begin{aligned} \text{var } n^{1/2} (\tilde{\theta}_i - \theta_i) &= \frac{1}{(A_i(r_n, c_n))^2} \text{var} \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \varphi(y - y_{ij}, r_n) dy \\ &\leq \frac{1}{(A_i(r_n, c_n))^2} E \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \varphi(y - y_{ij}, r_n) dy \\ &\leq \frac{1}{A_i(r_n, c_n)} \end{aligned}$$

donc $\text{var } n^{1/2} (\tilde{\theta}_i - \theta_i) \leq \frac{1}{I}$ mais $\tilde{\theta}_i$ ne peut être super efficace puisqu'équivariant

on a donc $\text{var } n^{1/2} (\tilde{\theta}_i - \theta_i) = \frac{1}{I}$

$$\text{var } n^{1/2} (\tilde{\alpha} - \alpha) = \frac{\sum_{i=1}^k n_i \sum_{j=1}^{n_i} X_{ij}^2 \text{var} \int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \varphi(y - y_{ij}, r_n) dy}{\left(\sum_{i=1}^k n_i b_i A_i(r_n, c_n)\right)^2}$$

(étant donné l'indépendance de l'échantillonnage) d'où $\text{var } n^{1/2} (\tilde{\alpha} - \alpha) = \frac{1}{\sum_{i=1}^k b_i I(f_i)}$

$$\begin{aligned} \text{cov } n^{1/2} (\tilde{\alpha} - \alpha) (\tilde{\theta}_i - \theta_i) &= \frac{n \sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} E \left(\int L_i(y, r_n) g\left(\frac{y}{c_n}\right) \varphi(y - y_{ij}, r_n) dy \right)^2}{\left(\sum_{i=1}^k n_i b_i A_i(r_n, c_n)\right) A_i(r_n, c_n)} \\ &= 0 \end{aligned}$$

IV - PARTIE ADAPTATIVE.

On obtient l'estimateur adaptatif $\hat{\theta}$ en remplaçant dans $\tilde{\theta}$ f_i par son estimation \hat{f}_{n_i} définie plus haut (partie 2).

$$\begin{aligned} \text{Soient : } \hat{L}_{n_i}(x, r) &= \frac{\hat{f}_{n_i}'(x, r)}{\hat{f}_{n_i}(x, r)} \\ \hat{A}_{n_i}(r, c) &= \int \hat{L}_{n_i}^2(x, r) g\left(\frac{x}{c}\right) \hat{f}_{n_i}(x, r) dx \\ \hat{\theta}_i &= \bar{\theta}_i - \frac{\int \hat{L}_i(y, r_n) g\left(\frac{y}{c_n}\right) \sum_{j=1}^{n_i} \frac{1}{n_i} \varphi(y - E_{ij} + \bar{\theta}_i + \bar{\alpha} X_{ij}, r_n)}{\hat{A}_{n_i}(r_n, c_n)} \\ &\quad + c_{\theta_i} \\ \hat{\alpha} &= \bar{\alpha} - \frac{\sum_{i=1}^k \int \hat{L}_i(y, r_n) g\left(\frac{y}{c_n}\right) \sum_{j=1}^{n_i} X_{ij} \varphi(y - E_{ij} + \bar{\theta}_i + \bar{\alpha} X_{ij}, r_n)}{\sum_{i=1}^k n_i b_i \hat{A}_{n_i}(r_n, c_n)} \end{aligned}$$

Etudions quelques propriétés des $\hat{f}_{n_i}(x, r_n)$. Des propriétés des fonctions $\psi_{n_i}(y, r_n, c_{ij})$ vues en annexes il vient le contrôle des dérivées de \hat{f}_{n_i} :

$$\hat{f}_{n_i}^{(\nu)}(x, r_n) = \frac{O_p(1)}{r_n^{\nu+\epsilon}} (\hat{f}_{n_i}(x, r_n) + \hat{f}_{n_i}^{1-\epsilon}(x, r_n)) \quad (13)$$

Si on pose $\psi_i(x, c_{ij}, r_n) = \frac{1}{n_i} \sum_{j=1}^{n_i} f_i(x - E_{ij} + \bar{\theta}_i + \bar{\alpha} X_{ij}, r_n)$ on a :

$$\begin{aligned} &\int (\psi_i^{(\nu)}(x, c_{ij}, r_n) + \psi_i^{(\nu)}(-x, c_{ij}, r_n) - 2 f_i^{(\nu)}(x, r_n))^2 \\ &\leq M \left(\sup_j |(\bar{\theta}_i - \theta_i) + (\bar{\alpha} - \alpha) X_{ij}| \right)^4 \int f_i^{(\nu+2)}(x, r_n)^2 dx \leq \frac{O_p(1)}{n^{1+\delta} r_n^{2\nu+5}} \end{aligned}$$

De ce résultat, de l'inégalité (8), annexe 2, et en utilisant la symétrie de $f_i(x, r_n)$ et $\hat{f}_{n_i}(x_n, r_n)$ on déduit le théorème suivant :

$$\text{si } \frac{1}{r_n^{4n}} = O_p(1), \quad \text{alors}$$

$$\begin{aligned} \int_{-\infty}^{+\infty} (\hat{f}_{n_i}^{(\nu)}(x, r_n) - f_i^{(\nu)}(x, r_n))^2 dx &= O_p(1) \left(\frac{1}{n^2 r_n^{2\nu+5}} + \frac{1}{n r_n^{2\nu+1}} \right) \\ &= \frac{O_p(1)}{n r_n^{2\nu+1}} \end{aligned} \quad (14)$$

De (13) on déduit

$$\hat{L}_{n_i}(x, r_n) = \frac{O_p(1)}{r_n^{1+\epsilon}} (1 + \hat{f}_{n_i}^{-\epsilon}(x, r_n)) \quad (15)$$

On a une inégalité du même type pour $L_i(x, r)$:

$$L_i(x, r_n) = \frac{O_p(1)}{r_n^{1+\epsilon}} (1 + f_i^{-\epsilon}(x, r_n)) \quad (16)$$

Ces propriétés mettent en valeur les similitudes de \hat{f}_{n_i} et f_i et étudient leurs distances.

Pour montrer que le modèle estimé est proche du modèle réel nous allons utiliser ces propriétés pour montrer la continuité de l'information. Pour cela on reprendra la démonstration faite par STONE dans [87] qui traite à part les petites valeurs de f_i et \hat{f}_{n_i} et utilise (14), (15), (16) pour démontrer le théorème :

Sous la condition : $\frac{c_n}{n^{1-\epsilon} r_n^\delta} = O_p(1)$ pour $\epsilon > 1 - \delta$ alors

$$\hat{A}_{n_i}(r_n, c_n) = A_i(r_n, c_n) + O_p(1)$$

Il reste à démontrer que l'estimateur adaptatif $\hat{\Theta}$ est asymptotiquement proche de $\tilde{\Theta}$ c'est-à-dire le théorème suivant :

Sous la condition $\frac{dn^2}{n^{1-\epsilon} t_n^6} = O(1)$ quand $n \rightarrow \infty$ alors $n^{1/2} (\tilde{\Theta} - \hat{\Theta})$ tend vers 0 en probabilité quand n tend vers l'infini.

Pour cette démonstration la symétrie de f_i (et \hat{f}_i) est essentielle. Pour ne pas rompre cette symétrie, on sera amené à choisir g symétrique autour de l'origine.

Pour démontrer ce théorème il suffit de s'intéresser aux quantités :

$$J = \int (\hat{L}_i(x, r_n) - L_i(x, r_n)) g\left(\frac{x}{c_n}\right) \left(\sum_{j=1}^{n_i} \frac{1}{n_i} \varphi(y - E_{ij} + \bar{\theta}_i + \bar{\alpha} X_{ij}) - f_i(x, r_n) \right)$$

pour les quantités $\tilde{\theta}_i - \hat{\theta}_i$

$$\text{et } J' = \int (\hat{L}_i(x, r_n) - L_i(x, r_n)) g\left(\frac{x}{c_n}\right) \sum_{j=1}^{n_i} \frac{X_{ij}}{n_i} (\varphi(y - E_{ij} + \bar{\theta}_i + \bar{\alpha} X_{ij}) - f_i(x, r_n))$$

Rappelons les notations :

$$\psi_{n_i}(x, c_{ij}, r_n) = \frac{1}{n_i} \sum_{j=1}^{n_i} \varphi(x - y_{ij} + c_{ij}; r_n)$$

$$\psi_{n_i}^*(x, c_{ij}, r_n) = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \varphi(x - y_{ij} + c_{ij}; r_n)$$

$$c_{ij} = (\bar{\alpha} - \alpha) X_{ij} + (\bar{\theta}_i - \theta_i)$$

Pour obtenir une majoration des intégrales J et J' on va partitionner le domaine d'intégration (pour isoler les petites valeurs de f_i et \hat{f}_{n_i}). On appellera $J_{\mathcal{B}}$ (respectivement $J'_{\mathcal{B}}$) l'intégrale J (respectivement J') restreinte au domaine \mathcal{B} .

(1) Sur le domaine $\mathcal{B}_1 = \{ \hat{f}_{n_i} < \frac{1}{2}, f_i(x, r_n) \geq \hat{f}_{n_i}(x, r_n) \}$ comme

$$\psi_{n_i}(x, c_{ij}, r_n) \leq 2 \hat{f}_{n_i}(x, r_n)$$

On a en utilisant (15), (16) et la symétrie :

$$J_{\mathcal{B}_1} = O_p(1) \frac{c_n}{n^{2-\epsilon} r_n}$$

Pour J' , il suffit de remarquer :

$$\begin{aligned} \psi_{n_i}^*(x, c_{ij}, r_n) &\leq \sup_j |X_{ij}| \psi_{n_i}(x, c_{ij}, r_n) \\ &\leq 2 M \hat{f}_{n_i} \frac{1-\delta}{4} \end{aligned}$$

$$\text{d'où} \quad J_{\mathcal{B}_1} = O_p(1) \frac{c_n}{n^{\frac{7+\delta}{4} - \epsilon} r_n}$$

(2) Sur le domaine $\mathcal{B}_2 = \{ f_i(x, r_n) < \frac{1}{n_i^2}, \hat{f}_{n_i}(x, r_n) > f_i(x, r_n) \}$.

Les propriétés (13), (15), (7), (10), annexe 2) entraînent :

$$J_{\mathcal{B}_2} = O_p(1) \frac{c_n}{n^{1-\epsilon} r_n^2}$$

$$J'_{\mathcal{B}_2} = O_p(1) \frac{c_n}{n^{1-\epsilon} r_n^2}$$

(3) Sur le domaine $\mathcal{B}_3 = \{ \hat{f}_{n_i}(x, r_n) > \frac{1}{n_i^2}, f_i(x, r_n) > \frac{1}{n_i^2} \}$ on peut écrire :

$$\hat{L}_{n_i}(x, r_n) - L_i(x, r_n) = \frac{\hat{f}'_{n_i}(x, r_n) - f'_i(x, r_n)}{f_i(x, r_n)} - \frac{\hat{f}'_{n_i}(x, r_n) (\hat{f}_{n_i}(x, r_n) - f_i(x, r_n))}{\hat{f}_{n_i}(x, r_n) f_i(x, r_n)}$$

Et on traitera séparément les deux membres obtenus .

Des propositions (1), (2) annexe 1 ; (11), (12) annexe 2 on déduit :

$$\int_{-c_n}^{+c_n} \frac{(\psi_{n_i}(x, r_n, c_{ij}) - f_i(x, r_n))^2}{f_i(x, r_n)} 1_{\{n_i^2 f_i(x, r_n) \geq 1\}} = \frac{O_p(1) c_n n^{\epsilon/2}}{n r_n^2}$$

$$\int_{-c_n}^{+c_n} \frac{(\psi'_{n_i}(x, r_n, c_{ij}) - f'_i(x, r_n))^2}{f_i(x, r_n)} 1_{\{n_i^2 f_i(x, r_n) \geq 1\}} = \frac{O_p(1) c_n n^{\epsilon/2}}{n r_n^3}$$

d'où

$$(17) \int_{-c_n}^{+c_n} \frac{(\hat{f}_{n_i}(x, r_n) - f_i(x, r_n))^2}{f_i(x, r_n)} 1_{\{n_i^2 f_i(x, r_n) \geq 1\}} = \frac{O_p(1) c_n n^{\epsilon/2}}{n r_n^2}$$

$$(18) \int_{-c_n}^{+c_n} \frac{(\hat{f}'_{n_i}(x, r_n) - f'_i(x, r_n))^2}{f_i(x, r_n)} 1_{\{n_i^2 f_i(x, r_n) \geq 1\}} = \frac{O_p(1) c_n n^{\epsilon/2}}{n r_n^3}$$

De l'inégalité de SCHWARTZ , (15), (17), (18), on déduit :

$$J_{\beta_3} = \frac{O_p(1) c_n}{n^{1-\epsilon} r_n^3}$$

$$J'_{\beta_3} = \frac{O_p(1) c_n}{n^{1-\epsilon} r_n^3}$$

Le théorème en résulte.

Annexe 1 :

Cette annexe a pour but de généraliser les propriétés des estimations régularisées, non symétrisées, des f_i (les fonctions $f_{n_i}(x, r_n) = \frac{1}{n_i} \sum_{j=1}^{n_i} \varphi(x - y_{ij})$) étudiées par STONE dans [87] à des fonctions

$$\psi_{n_i}(x, c_{ij}, r_n) = \sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} \varphi(x - y_{ij} + c_{ij}, r_n)$$

Ces propriétés sont utiles aux démonstrations principales et il y est fait référence.

On s'intéresse la plupart du temps à des majorations. On supposera donc $\gamma_{ij} \geq 0$ et $\frac{\sum_{j=1}^{n_i} \gamma_{ij}}{n} \leq M$. Dans cette annexe on s'intéresse à des propriétés locales des $\psi_{n_i}(x, c_{ij}, r_n)$.

1. Contrôle des dérivées :

On sait (cf. STONE [87]) qu'il existe δ_0 tel que pour $0 < \delta \leq \delta_0$ on a

$$|\varphi^{(\nu)}(x, r_n)| \leq 2 r_n^{-\nu} \delta^{-\epsilon} (\varphi(x, r_n) + \delta^2 r_n^{-1})$$

et en prenant :

$$\delta = \min \left(r_n, \frac{\psi_{n_i}(x, c_{ij}, r_n)}{\sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i}} \right)$$

on a :

$$(1) \quad |\psi_{n_i}^{(\nu)}(x, c_{ij}, r_n)| = O_p(1) r_n^{-(\nu+\epsilon)} (\psi_{n_i}(x, c_{ij}, r_n) + \psi_{n_i}^{1-\epsilon}(x, c_{ij}, r_n))$$

(Le signe de dérivation sous-entend dérivation par rapport à x).

2. Contrôle local sur un intervalle d'ordre $n^{-1/4}$.

On a pour $|x| < a_n \text{Log } L$

$$a_n = r_n^2 n^{1/4} M^{-1} \quad L > 1 :$$

$$\sup_{|t| < M n^{-1/4}} \varphi(x + t, r_n) \leq L \varphi(x, r_n)$$

$$\text{donc } \sup_{|c_{ij}| \leq M n^{-1/4} \quad \forall j, \forall i} \sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} \varphi(y - y_{ij} + c_{ij}, r_n) \leq L \sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} \varphi(y - y_{ij}, r_n) \\ + \sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} L^{-1/\epsilon}$$

(Puisque pour a fixé, $L_0 > 1$, $n > N$ on a sur un ensemble de mesure supérieure à $1 - a$:

$$\varphi(x + t - y, r_n) \leq L^{-1/\epsilon} \quad \text{pour} \quad \begin{cases} |x - y| \geq a_n \text{Log } L \\ |t| \leq M n^{-1/4} \\ L \geq L_0 \end{cases}$$

On vérifie donc pour $\epsilon > 0$ $M > 0$ sous la condition $\frac{1}{n r_n^3} = O_p(1)$.

$$(2) \sup_{|c_{ij}| \leq M n^{-1/4}, \forall j} \psi_{n_i}(y, c_{ij}, r_n) = O_p(1) (\psi_{n_i}(y, r_n) + \psi_{n_i}^{1-\epsilon}(y, r_n))$$

$$\text{où} \quad \psi_{n_i}(y, r_n) = \sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} \varphi(y - y_{ij}; r_n)$$

Annexe 2 :

On va s'intéresser maintenant aux propriétés des $\psi_{n_i}(y, r_n)$.

$\psi_{n_i}(y, r_n)$ est la densité obtenue en convoluant le noyau régularisant $\varphi(\cdot, r_n)$ avec la mesure μ_i définie par

$$\mu_i([-\infty, y]) = \sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} 1_{(y_{ij} < y)}$$

En appliquant HÖLDER on trouve les inégalités pour $q < r$

$$(3) \quad \frac{q}{r} \psi_{n_i}(y, q) \leq \psi_{n_i}(y, r) \leq \frac{q}{r} \left(\sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} \right)^{\frac{1}{1-\frac{q^2}{r^2}}} \psi_{n_i}(y, q)$$

ce qui traduit la continuité de la convolution par rapport au paramètre r .

Toujours en appliquant HÖLDER, on trouve

$$(4) \quad \int_{-c}^{+c} \psi_{n_i}(y, r_n)^{1-\epsilon} \leq \left(\sum_{j=1}^{n_i} \frac{\gamma_{ij}}{n_i} \right)^{1-\epsilon} (2c)^\epsilon \leq M (2c)^\epsilon \quad \text{pour } 0 < \epsilon < 1$$

$$\begin{aligned} \text{var } \psi_{n_i}^{(\nu)}(y, r_n) &= \sum_{j=1}^{n_i} \frac{\gamma_{ij}^2}{n_i^2} \text{var } \varphi^{(\nu)}(y - y_{ij}, r_n) \\ &\leq \sum_{j=1}^{n_i} \frac{\gamma_{ij}^2}{n_i^2} E \varphi^{(\nu)}(y - y_{ij}, r_n)^2 \end{aligned}$$

d'où

$$(5) \quad E [\psi_{n_i}^{(\nu)}(y, r_n) - \psi_i^{(\nu)}(y, r_n)]^2 \leq \frac{K_\nu}{r^{2\nu+1}} \sum_{j=1}^{n_i} \frac{\gamma_{ij}^2}{n_i^2} f_i(x, r_n)$$

$$\text{où} \quad \psi_i(x, r_n) = \frac{1}{n_i} \sum_{j=1}^{n_i} \gamma_{ij} f_i(x, r_n)$$

$$\text{On en déduit} \quad E \int_{-c}^{+c} \frac{\psi_{n_i}^2(y, r_n)}{f_i(y, r_n)} \leq \frac{2c}{r} \sum_{j=1}^{n_i} \frac{\gamma_{ij}^2}{n_i^2} + \left(\frac{\sum_{j=1}^{n_i} \gamma_{ij}}{n_i} \right)^2$$

et en appliquant HÖLDER

$$(6) \quad E \int_{-c}^{+c} \frac{\psi_{n_i}^{(1-\beta)}(y, r_n)}{f_i(y, r_n)^\epsilon} \leq \left(\frac{\sum_{j=1}^{n_i} \gamma_{ij}}{n_i} \right)^{1-\epsilon-\beta} (2c)^{3+\epsilon} \left[\frac{2c}{r} \frac{\sum_{j=1}^{n_i} \gamma_{ij}^2}{n_i^2} + \left(\frac{\sum_{j=1}^{n_i} \gamma_{ij}}{n_i} \right)^2 \right]$$

$$\text{si } 2\epsilon + \beta < 1$$

$$\text{d'où si } \frac{c_n}{n^\delta r_n} = O_p(1) \quad \text{et} \quad \frac{\sum \gamma_{ij}^2}{n^2} \leq \frac{M}{n^\delta} \quad \delta > 0 :$$

$$(7) \quad \int_{-c_n}^{+c_n} \frac{\psi_{n_i}(x, r_n)^{1-\beta}}{f_i^\epsilon(x, r_n)} = O_p(1) n^{2(\beta + \epsilon)}$$

Ces majorations sont utiles pour démontrer la convergence de l'estimateur $\tilde{\Theta}$.

Il y a alors à donner à (γ_{ij}) les valeurs successives $\gamma_{ij} = 1 \quad \forall j$, $\gamma_{ij} = |X_{ij}|$, $\gamma_{ij} = X_{ij}^2$, $\gamma_{ij} = |X_{ij}|^3$. On peut vérifier qu'avec les conditions imposées aux X_{ij} on a toujours $\frac{\sum \gamma_{ij}^2}{n^2} \leq \frac{M}{n^\delta}$ et $\frac{\sum \gamma_{ij}}{n} \leq M$.

On peut déduire de (5) que si $\frac{\sum \gamma_{ij}^2}{n^2} \leq M n^{-\delta} \quad \delta > 0$ alors

$$(8) \quad \int_{-\infty}^{+\infty} (\psi_{n_i}^{(\nu)}(x, r_n) - \psi_i^{(\nu)}(x, r_n))^2 dx = \frac{O_p(1)}{n^\delta r_n^{2\nu+1}} \quad \text{et}$$

$$(9) \quad \int_{-c_n}^{+c_n} \frac{(\psi_{n_i}^{(\nu)}(x, r_n) - \psi_i^{(\nu)}(x, r_n))^2}{f_i^\epsilon(x, r_n)} = \frac{O_p(1) n^{2\epsilon}}{n^\delta r_n^{2\nu+1}}$$

$$\epsilon > 0$$

Ces égalités appliquées à $\psi_{n_i}(x, r_n) = \sum_{j=1}^{n_i} \frac{1}{n_i} \varphi(y - y_{ij}, r_n)$ évaluent une distance entre f_i et son estimation.

Cette partie sera utilisée dans la partie adaptative pour montrer que le modèle estimé est proche du modèle réel (au sens $n^{1/2}(\tilde{\Theta} - \hat{\Theta}) \rightarrow 0$ en probabilité).

Prenant $\gamma_{ij} = 1 \quad \forall j$ on pourra supposer $\delta = 1$. $(\sum_{j=1}^{n_i} \frac{\gamma_{ij}^2}{n_i^2} = \frac{1}{n_i})$.

$$\text{Alors } E \int_{-c}^{+c} \frac{\psi_{n_i}(x, r)}{f_i(x, r)} 1_{\{f_i(x, r) < \frac{1}{n^\gamma}\}} \leq 2c \left(\frac{1}{nr} + \frac{1}{n^\gamma} \right)$$

d'où par HÖLDER :

$$E \int_{-c}^{+c} \frac{\psi_{n_i}^{1-\beta}(x, r)}{f_i^\epsilon(x, r)} 1_{\{f_i(x, r) < \frac{1}{n^\gamma}\}} \leq \frac{2c}{n^\gamma(1-\epsilon-\beta)} \left(1 + \frac{n^\gamma}{nr} \right)$$

$$\text{et si } \frac{c_n}{n r_n} = O_p(1)$$

$$(10) \quad \int_{-c_n}^{+c_n} \frac{\psi_{n_i}^{1-\beta}(x, r_n)}{f_i^\epsilon(x, r_n)} 1_{\{n^2 f_i(x, r_n) < 1\}} = O_p(1) \frac{c_n n^{2(\epsilon+\beta)}}{n r_n}$$

de même, on peut écrire :

$$(11) \quad \int_{-c_n}^{+c_n} \frac{(\psi_{n_i}^{(\nu)}(x, r_n) - \psi_i^{(\nu)}(x, r_n))^2}{f_i(x, r_n)} 1_{\{f_i(x, r_n) > \frac{1}{n^2}\}} = O_p(1) \frac{c_n n^\epsilon}{n r_n^{2\nu+1}}$$

Avec des raisonnements analogues (HÖLDER) on montre : pour $0 \leq \epsilon \leq 2$, $\beta \geq 0$,

$\gamma \geq 0$

$$E \int_{-c}^{+c} \frac{\psi_{n_i}^{2-\epsilon}(x, r)}{f_i^{1+\beta}(x, r)} 1_{\{n^\gamma f_i(x, r) \geq 1\}} \leq n^{\gamma\beta + \gamma\epsilon/2} (2c)^{\epsilon/2} \left(1 + \frac{2c}{nr}\right)$$

d'où si $\frac{c_n}{n r_n} = O_p(1)$ alors

$$(12) \quad \int_{-c_n}^{+c_n} \frac{\psi_{n_i}^{2-\epsilon}(x, r_n)}{f_i(x, r_n)} 1_{\{n^2 f_i(x, r_n) \geq 1\}} dx = O_p(1) n^{2\epsilon}.$$

Dominique PICARD

Mathématiques

Université de Paris-Sud

91405 ORSAY.

BIBLIOGRAPHIE

- [1] ANDREWS F.C. - Asymptotic behavior of some rank tests for analysis of variance. *Ann. Math. Stat.* 25, pp. 724-735 (1954).
- [2] ANDREWS D.F. - (BICKEL P.J. - HAMPEL F.R. - HUBER P.J. - ROGERS W.H. - TUKEY J.W.) - Robust Estimates of location (Survey and Advances), Princeton University Press, 1972, 373 p.
- [3] ANSCOMBE F.J. - Rejection of outliers - *Technometrics* 2, pp. 123-147, (1960).
- [4] AZENCOTT R. - Ce même volume, Chapitre n° V.
- [5] BENNET J. - Asymptotic properties of ideal linear estimates - Dissertation (Michigan), 1952.
- [6] BICKEL P.J. - Some contributions to the theory of order statistics - *Proc. 5th Berkeley Symp.* pp. 575-591 (1967).
- [7] BICKEL P.J., HODGES J.L. - The asymptotic theory of Galton test and a related simple estimate location - *Ann. Math. Stat.* 38, pp. 73-89 (1967).
- [8] BICKEL P.J., LEHMANN E.L. - Descriptive statistics for non parametric models - *Ann. Math. Stat.*, 3, n° 5 (1975).
- [9] BICKEL P.J. - On some Robust Estimates of location - *AMS* 36 (1965).
- [10] BILLINGSLEY P. - Convergence of probability measures - Wiley and son, New York (1968).
- [11] BLACKWELL D. - Comparaison of experiments - *Proc. of the 2nd Berkeley Symposium*, p. 93-102 (1951).
- [12] BHATTACHARYA P.K. - Efficient estimation of a shift parameter from group data - *Ann. Math. Stat.* 38, pp. 1770-1787 (1967).
- [13] BOURBAKI N. - Topologie générale - Chap. 4 (Hermann - Paris).
- [14] BOURBAKI N. - Topologie générale - Chap. 9 (Hermann - Paris).
- [15] BOURBAKI N. - Espaces vectoriels topologiques - Chap. 2 (Hermann-Paris).
- [16] CHERNOFF H., GASTWIRTH J.L., JOHNS M.U., Jr. - Asymptotic distributions of linear combinations of functions of order statistics with applications to estimation - *Ann. Math. Stat.* 38, pp. 52-72, (1967).

- [17] CHERNOFF H., SAVAGE R. - Asymptotic normality and efficiency of certain non parametric tests - Ann. Math Stat. 29, pp. 972-994 (1958).
- [18] CHOQUET G. - Theory of capacities - Ann. Inst. Fourier, n° 5, pp. 131-292 (1953-54).
- [19] CHOQUET G. - Forme abstraite du théorème de capacitable - Ann. Inst. Fourier, n° 9, pp. 83-89 (1959).
- [20] COURRÈGE P., PRIOURET P. - Sur l'évolution des intégrales multiples exprimant le niveau des tests F en statistique multinormale - Bull. Soc. Math. 95, pp. 93-150 (1971).
- [21] CRAMER H. - Mathematical methods of Statistics (1946).
- [22] CUTHBERT D. - Locating outliers in factorial experiments - Technometrics 2, pp. 149-156 (1960).
- [23] DELLACHERIE C. - Capacités et processus stochastiques (Springer - Verlin) (1972).
- [24] DENIAU C., VIANO C., OPPENHEIM G. - Ce même volume, Chapitre n° X.
- [25] DIEUDONNÉ J. - Eléments d'Analyse II, Hermann, (1968).
- [26] DUNFORD N., SCHWARTZ J.T. - Linear operators - Vol. 1 (Wiley, New York) (1958).
- [27] DIXON W.J., TUKEY J.W. - Approximate behavior of the distribution of Winsorised (Trimming/Winsorisation 2) - Technometrics 10, pp. 83-98, (1968).
- [28] FILIPPOVA A.A. - Mises' theorem on the asymptotic behavior of empirical distribution functions and its statistical applications - Th. of Prob. and its app. 7, pp. 24-57 (1962).
- [29] FOURGEAUD C., FUCHS A. - Statistique, pp. 212-215 - Collection Universitaire de Mathématiques, Dunod éd. (1967).
- [30] FREEDMAN H.W. - The "little variable factor". A statistical discussion of the reading of seismograms - Bull. of the Seismol. Soc. of America 56, pp. 593-604 (1966).
- [31] HAJEK J., SIDAK Z. - Theory of Rank-Test, Academic Press, N.Y. (1967).
- [32] HAJEK J. - Non parametric statistics - Ed. Holden Day, San Francisco, Californie (1969).
- [33] HAJEK J. - Local asymptotic minimax and admissibilities in estimation - Proc. Sixth Berkeley Symp. Math. Stat. Prob. (1972).
- [34] HAMPEL F.R. - Contributions to the theory of robust estimation - Univ. of California, Berkeley (Thèse, 1968).

- [35] HAMPEL F.R. - A general qualitative definition of robustness - Ann. Math. Stat. 42, n° 6, pp. 1887-1896 (1971).
- [36] HAMPEL F.R. - The influence curve and its role in robust estimation - J.A.S.A., 69, n° 346, pp. 383-393 (1974).
- [37] HODGES J.L. - Galton's rank order test - Biometrika, 42, p. 261-262 (1955).
- [38] HODGES J.L. - Efficiency in normal samples and tolerance of extreme values. Proc. 5th Berk. Symp. pp. 163-186 (1967).
- [39] HODGES J.L. - Efficiency in normal samples and tolerance of extreme values for some estimates of location - 5th Berkeley Symposium in Probability - vol. 1, p. 163-186 (1966).
- [40] HODGES J.L., LEHMANN E. L. - Efficiency of some non parametric competitors of the t-test, Ann. Math. Stat. 27, p. 324-335 (1956).
- [41] HODGES J.L., LEHMANN E.L. - Estimates of location based on rank tests - Ann. Math. Stat. 34, p. 598-611 (1963).
- [42] HOEFFDING W., ROSENBLATT J.R. - The efficiency of tests - Ann. Math. Stat., pp. 52-63.
- [43] HOLLANDER M. et al. - Robustness of the Wilcoxon test to a certain dependence between samples - AS 2, n° 1, pp. 177-181 (1974).
- [44] HUBER P.J. - Robust estimation of a location parameter. Ann. Math. Stat. 35, 73-101 (1964).
- [45] HUBER P.J. - A robust version of the probability ratio test. Ann. Math. Stat. n° 36, p. 1753-1758 (1965).
- [46] HUBER P.J. - Théorie de l'inférence statistique robuste - Séminaire de mathématiques supérieures - Université de Montréal.
- [47] HUBER P.J. - The behavior of maximum likelihood estimates under non standard condition - Proc. Fifts Berk. Symp. on Math. Stat. and Prob., vol. 1, pp. 221-233 - (1967).
- [48] HUBER P.J. - Studentizing robust estimates in Nonparametric Techniques in Statistical Inference, Cambridge Univer. Press, 453-463 (1970).
- [48 bis] HUBER P.J. - Robust confidence limits. Z.Wahrscheinlichkeisteorie verw. Geb. 10, pp. 269-278 (1968).
- [49] HUBER P.J. - The 1972 Wald Lecture robust statistics : a review - Ann. Math. Stat. 43 p. 1041-1067 (1972).
- [50] HUBER P.J. - Robust regression asymptotic conjectures and Monte Carlo - Ann. Math. Stat. vol. 1, n° 5 (1973).

- [51] HUBER P.J., STRASSEN V. - Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. of Stat.*, vol. 1, n° 2, p. 251-263 (1973).
- [52] HUBER P.J. - Fisher information and spline interpolation. *The annals of Statistics* 2, p. 1029-1034 (1974).
- [53] HUBER P.J., STRASSEN V. - Correction to minimax tests and the Neyman-Pearson lemma for capacities - *Ann. of Stat.* vol. 2, n° 1, p. 223-224 (1974).
- [54] HUBER C. - Etude asymptotique de tests robustes. Thèse à l'Ecole Polytechnique Fédérale de Zürich (1970).
- [55] HUBER C. - Ce même volume, Chapitre n° II.
- [56] HUBER C. - Ce même volume, Chapitre n° V.
- [57] HUBER C. - Ce même volume, Chapitre n° XII.
- [58] JAECKEL L.A. - Robust estimates of location (thèse), Berkeley (1969).
- [59] JAECKEL L.A. - Some flexible estimates of location - *Ann. Math. Stat.* 42, n° 5, pp. 1540-1552 (1971).
- [60] KALLIANPUR G. - Von Mises functionals and maximum likelihood estimation - *Contributions to Statistics*, pp. 137-146 (1963).
- [61] KEMPTHORNE O. - Some aspects of experimental inference, *J.A.S.A.* 61, pp. 11-34 (1966).
- [62] KENDALL M.G. - The advanced theory of statistics, vol. II, Ed. Ch. Griffin and Co Ltd, London (1967).
- [63] KLOTZ J. - Small sample power and efficiency for the one-sample Wilcoxon and normal scores tests - *Ann. Math. Stat.* 34, pp. 624-632, (1963).
- [64] LE CAM L. - Locally asymptotically normal families of distributions, *Univ. of Calif. Publ. in Stat.* 3 (1960).
- [65] LE CAM L. - Sufficiency and approximate sufficiency - *Ann. Math. Stat.* vol. 35, p. 1419-1455 (1964).
- [66] LE CAM L. - *Ann. Math. Stat.* 41, p. 802-828 (1970).
- [67] LE CAM L. - Notes on asymptotic methods in statistical decision theory - Centre de Recherches Mathématiques - Université de Montréal.
- [68] LEHMANN E.L. - Testing statistical hypotheses, Wiley, N.Y. (1959).
- [69] LINNIK Ju.V. - Statistical problems with nuisance parameters - *Trans. Am. Maths. Soc.*, vol. 20 (1968).

- [70] LOÈVE M. - Probability theory, Van Nostrand, (1955).
- [71] MEYER P.A. - Probabilités et potentiel (Hermann-Paris) (1966).
- [72] MOORE D.S. - An elementary proof of asymptotic normality of linear functions of order statistics, *Ann. Math. Stat.* 39, n° 1, pp. 263-265 (1968).
- [73] NEVEU J. - Bases mathématiques du calcul des probabilités (Masson, Paris) (1964).
- [74] NOETHER G.E. - On a theorem of Pitman, *Ann. Math. Stat.* 26, pp. 64-68, (1955).
- [75] PFANZAGL J. - Preprints in Statistics - Université de Cologne, n° 7 (1974).
- [76] PROKHOROV Yu. V. - Convergence of random processes and limit theorem in probability theory - Theory of Prob. and applications, vol. 1, pp. 157-214 (1956).
- [77] PYKE R., SHORACK G.R. - Weak convergence of the two sample empirical process and a new approach to Chernoff-Savage theorems, *Ann. Math. Stat.*, pp. 755-771 (1968).
- [78] RELLES D.A. - Robust regression by modified least squares. Ph. D. Yale University (1967).
- [79] RELLES D.A. - Variance reduction techniques for Monte Carlo sampling from Student distributions - *Technometrics*, 12, 499-516 (1970).
- [80] SCHEFFE H. - The analysis of variance (Wiley).
- [81] SCHOLZ F.W. - A comparison of efficient location estimations - *Annals of Statistics*, p. 1323-1326, vol. 2 (1974).
- [82] SHORACK G.R. - Functions of order statistics - *Ann. Math. Stat.* 43, n° 2, pp. 412-427 (1972).
- [83] SHORACK G.R. - Random means, *AS.* 2, n° 4, pp. 661-675 (1974).
- [84] SKOROHOD A.V. - Limit theorems for stochastic processes - Theory of prob. and applications (Translated by SIAM), 1, pp. 261-290 (1956).
- [85] STEIN C. - Proceedings of the third Berkeley Symposium on Math. Stat. and Probability. Efficient nonparametric Testing and Estimation pp. 187-195 (1956).
- [86] STONE C.J. - Asymptotic properties of estimations of a location parameter - *Annals of Stat.*, 2, pp. 1127-1137 (1974).
- [87] STONE C.J. - Adaptative maximum likelihood of a location parameter - *Annals of Statistics*, 3, pp. 267-284 (1975).

- [88] STRASSEN V. - Messfehler und Information - Z. für Wahrscheinlichkeitstheorie und Verw. Gebiete, vol. 2, p. 273-305 (1964).
- [89] STIGLER J., STEPHEN M.-Simon Newcombe, Percy Daniell, and the history of robust estimation 1885-1920. Unpublished report of the Univ. of Wisconsin, Dept. of Statistics.
- [90] STRASSEN V. - The existence of probability measures with given marginals - Ann. Math. Statist., 36, p. 423-439 (1965).
- [91] TORGERSEN E.N. - Comparison of experiments when the parameter space is finite. Z. für Wahrscheinlichkeitstheorie und Verw. Gebiete, vol. 16, p. 219-249 (1970).
- [92] TUKEY J.W. - Which part of the sample contains the information ? - Proc. of the Nat. Acad. of Sciences, 53, p. 127-134 (1956).
- [93] TUKEY J.W. - A survey of sampling from contaminated distributions - Contrib. to Prob. and Stat. (ed. Olkin), pp. 448-485, Stanford University Press (1960).
- [94] VAN EEDEN C. - Nonparametric estimation. 7th Session of the "Séminaire de Math. Sup". Presses de l'Univ. de Montréal (1968).
- [95] VAN EEDEN C. - Efficient robust estimation of location - Ann. Math. Stat., 41, pp. 172-181 (1970).
- [96] VAN EEDEN C. - Maximum likelihood estimation of partially or completely ordered parameters - Proc. Kon. Ned. Akad. v. Wet. A 60 (1957), 128-136, 201-211.
- [97] VON MISES R. - On the asymptotic distributions of differentiable statistical functions, Ann. Math. Stat. 18, pp. 309-348 (1947).
- [98] VON MISES R. - Mathematical theory of probability and statistics. Academic Press (1964).
- [99] YVILSAKER D., SACKS J. - A note on Huber's robust estimation of a location parameter, Ann. Math. Stat., 43, p. 1068-1075 (1972).
- [100] YOHAJ V.J. - Robust estimation in the linear model - Annals of Stat., 2, pp. 562-567, (1974).
- [101] ZACKS S. - The theory of statistical inference (Wiley), (1971).