

Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

1215

Lectures in Probability and Statistics

Lectures given at the Winter School in
Probability and Statistics
held in Santiago de Chile

Edited by G. del Pino and R. Rebolledo



Springer-Verlag

Berlin Heidelberg New York London Paris Tokyo

Editors

Guido del Pino

Rolando Rebolledo

Facultad de Matemáticas, Pontificia Universidad Católica de Chile

Casilla 6177, Santiago de Chile, Chile

Authors

Catherine Huber

Laboratoire de Statistique Médicale, Université de Paris V

45, rue des Saints Pères, 75270 Paris Cedex 06, France

Hermann Rost

Institut für Angewandte Mathematik, Universität Heidelberg

Im Neuenheimer Feld 294, 6900 Heidelberg, Federal Republic of Germany

Victor Solo

Department of Statistics, Harvard University

Cambridge, MA 02138, USA

John B. Walsh

Department of Mathematics, University of British Columbia

Vancouver (8), British Columbia, Canada

Mathematics Subject Classification (1980): 60K35; 60K25; 60G44; 60H05;
62F35; 62M10

ISBN 3-540-16822-2 Springer-Verlag Berlin Heidelberg New York

ISBN 0-387-16822-2 Springer-Verlag New York Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically those of translation, reprinting, re-use of illustrations, broadcasting, reproduction by photocopying machine or similar means, and storage in data banks. Under § 54 of the German Copyright Law where copies are made for other than private use, a fee is payable to "Verwertungsgesellschaft Wort", Munich.

© Springer-Verlag Berlin Heidelberg 1986

Printed in Germany

Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.

2146/3140-543210

INTRODUCTION

Ce recueil contient une partie des cours qui ont été faits à Santiago du Chili, dans le cadre de l'Ecole d'Hiver de Probabilités et Statistiques.

L'Ecole d'Hiver - qui en est déjà à sa cinquième année de vie - est devenue un lien de rencontre et d'échanges très nécessaire pour les chercheurs latino-américains que nous espérons perfectionner et élargir au cours du temps.

Nous remercions les auteurs du volume, Madame HUBER et Messieurs ROST, SOLO, WALSH, pour l'excellent matériel et le soin dédié aux cours, ainsi que tous les participants aux Ecoles pour l'intérêt porté à l'initiative.

Finalement, l'Ecole d'Hiver a pu avoir lieu grâce à la collaboration de la Direction de Recherche de l'Université Catholique du Chili (DIUC), du CIENES (Centre Interaméricain d'Enseignement de la Statistique); de la Coopération Scientifique Franco-Chilienne; de IBM-Chili; nous tenons à exprimer nos remerciements à toutes ces institutions.

GUIDO DEL PINO - ROLANDO REBOLLEDO
Professeurs à l'Université Catholique du Chili
Casilla 6177. Santiago
Chile

TABLE OF CONTENTS

CATHERINE HUBER	
Théorie de la robustesse	1
HERMANN ROST	
On the Behaviour of the Hydrodynamical Limit for Stochastic Particle Systems	129
VICTOR SOLO	
Topics in Advanced Time Series Analysis	165
JOHN B. WALSH	
✓ Martingales With a Multidimensional Parameter and Stochastic Integrals in the Plane	329

THEORIE DE LA ROBUSTESSE

Catherine Huber

INTRODUCTION

Du souci de se libérer d'hypothèses trop restrictives sur les modèles, sont nées diverses techniques d'inférence statistique qui peuvent être regroupées sous le terme de méthodes robustes. Ce qualificatif provient de ce que ces méthodes se défendent bien contre l'instabilité qui caractérise les procédures optimales pour certains modèles paramétriques et en particulier pour les modèles gaussiens. Ce phénomène d'instabilité, connu depuis longtemps, a été dénoncé, pour la première fois de manière systématique, par J.W. TUKEY en 1960. Des exemples en sont donnés au chapitre I, dont certains sont très classiques et d'autres moins, comme ceux relatifs aux tests de rangs.

C'est P.J. HUBER qui, le premier, a dans les années soixante développé une théorie cohérente de la statistique robuste, en commençant par stabiliser les estimateurs du maximum de vraisemblance par plongement dans une classe plus générale, celle des M-estimateurs. Des estimateurs stables ont été ensuite construits dans diverses classes, vérifiant une propriété d'optimalité de type minimax pour un modèle élargi par rapport au modèle initial. C'est l'objet du chapitre III.

Cependant, l'élargissement des hypothèses de travail peut se faire de deux manières essentiellement différentes selon que l'on privilégie l'un ou l'autre des deux éléments du couple formé par le modèle et le problème d'inférence statistique correspondant. Un exemple illustre assez bien la différence entre les deux points de vue : supposons que l'on veuille estimer la moyenne θ d'une variable normale de variance 1 dont on connaît n réalisations indépendantes. La moyenne des observations est un bon estimateur de θ qui devient mauvais si l'hypothèse de normalité n'est pas vérifiée. Pour élargir le modèle initial, qu'on peut noter $H = \{N(\theta, 1) ; \theta \in \mathbb{R}\}$, on peut choisir l'ensemble des probabilités proches de H pour une distance d sur l'ensemble des probabilités, soit $V(H) = \{P : \exists \theta \in \mathbb{R} : d(P, N(\theta, 1)) \leq \epsilon\}$ pour un ϵ petit. C'est, en gros, le point de vue de P.J. HUBER qui détermine à partir de là des estimateurs optimaux pour un critère minimax. Le second point de vue consiste à considérer le problème posé par l'estimation de θ comme un cas particulier du problème général de l'estimation du centre de symétrie d'une loi symétrique. L'hypothèse de normalité est ainsi abandonnée

au profit de l'hypothèse beaucoup plus générale de symétrie. C'est le point de vue adopté par C. STONE qui construit un estimateur optimal pour ce modèle général.

Ainsi l'élargissement des hypothèses de travail peut se faire, soit en tolérant de faibles écarts par rapport à un modèle paramétrique central, le problème d'inférence restant toujours relatif à ce modèle central, soit en plongeant le couple que forment le modèle et le problème associé dans un ensemble de tels couples et en résolvant le problème ainsi généralisé.

Ces deux points de vue correspondent à des situations concrètes très différentes, le premier supposant une connaissance assez précise du modèle, ce qui peut se rencontrer en physique par exemple, le second supposant que l'on dispose d'une assez grande quantité d'observations, et relèvent par suite, de méthodes différentes. Ils sont présentés en détail dans le chapitre II. Il paraît assez clair que, si le second point de vue est assez satisfaisant du point de vue mathématique, le premier est nettement plus réaliste.

Le premier point de vue, qui est celui considéré par P.J. HUBER et F.R. HAMPEL, conduit à des critères de type minimax et utilise essentiellement des arguments de convexité. Le second point de vue, successivement étudié par C. STONE, P.W. MILLAR, P.J. BICKEL et R. BERAN, s'oriente vers des méthodes de type non paramétrique incluant des estimations fonctionnelles (quelques aspects de ces estimations fonctionnelles sont présentés dans le dernier chapitre) ; il fait l'objet des chapitres IV et V. En ce qui concerne ce deuxième point de vue, lorsque le nombre des observations est suffisamment important, un moyen d'atteindre l'optimalité, malgré la grande taille du modèle, est l'adaptativité, introduite par C. STEIN et développé ensuite par P.J. BICKEL. Les méthodes adaptatives font l'objet du chapitre VI.

A la suite des chapitres III à VI qui, tous, concernent l'estimation, vient le chapitre VII consacré aux tests robustes qui consistent en des généralisations du lemme de NEYMAN et PEARSON, initialisées par P.J. HUBER en 1965 puis étudiées successivement par H. RIEDER pour les distances du type de celle de PROKHOROV et par L. BIRGE, pour celle de HELLINGER, qui développe des idées introduites par L. LECAM dans le cadre paramétrique.

Certains problèmes d'estimation robuste multidimensionnelle tels que l'estimation robuste de matrices de covariance ne sont pas traités ici. On peut les trouver dans le livre de P.J. HUBER "Robust Statistics" (1981).

TABLE DES MATIERES

CHAPITRE :

I - EXEMPLES D'INSTABILITE DES PERFORMANCES ET TESTS ET ESTIMATEURS USUELS	6
II - EVOLUTION DES POINTS DE VUE SUR LA ROBUSTESSE	19
III - A - ESTIMATEURS DU MAXIMUM DE VRAISEMBLANCE GENERALISE (M-estimateurs)	34
B - ESTIMATEURS FONDES SUR LES COMBINAISONS LINEAIRES DE STATISTIQUES D'ORDRE	49
C - ESTIMATEURS FONDES SUR LES RANGS (R-estimateurs) ...	53
IV - ESTIMATION ROBUSTE OPTIMALE PAR LA METHODE DU PLUS PROCHE VOISIN	60
V - ESTIMATION ROBUSTE OPTIMALE PAR LA METHODE DU MAXIMUM DE VRAISEMBLANCE AU PREMIER PAS : (OU PLUS PROCHE VOISIN POUR LA DISTANCE DE HELLINGER)	70
VI - ADAPTATIVITE	75
VII - TESTS ROBUSTES	95
VIII - ESTIMATION FONCTIONNELLE - RISQUE MINIMAX	118

I - EXEMPLES D'INSTABILITE DES PERFORMANCES DE
TESTS ET ESTIMATEURS USUELS

1 - LA MOYENNE EMPIRIQUE ET LES LOIS CONTAMINEES

C'est l'un des plus classiques et des plus anciens exemples de l'importance de la perturbation du comportement d'un estimateur provoquée par un léger écart de la loi des observations par rapport au modèle supposé. Il est dû à Tukey (1960) et c'est à partir de là que s'est développée la théorie de la robustesse.

On sait, que dans le modèle de translation $F^n(\cdot - \theta)$, $\theta \in \mathbb{R}$, lorsque F est la fonction de répartition de la loi normale sur \mathbb{R} , $N(0, \sigma^2)$, l'estimateur optimal (efficace) de θ est la moyenne empirique \bar{X}_n . Mais, si F est remplacé par $F_\varepsilon = (1-\varepsilon)N(\theta, \sigma^2) + \varepsilon N(\theta, 9\sigma^2)$, pour une petite valeur de ε n'excédant pas 0,1, l'efficacité de \bar{X}_n diminue beaucoup même si cette contamination par une loi plus dispersée n'est pratiquement pas décelable sur l'échantillon observé.

En effet, si la fraction de contamination est de 1% et la taille de l'échantillon $n = 1000$, 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 presque. Or il y en aura en moyenne 40% : En confondant, chaque fois qu'il n'y a pas de confusion possible, fonction de répartition et probabilité correspondante, pour $\varepsilon = 0,01$

$$F_\varepsilon(|X| > 2,5 \sigma) = F(|X| > \frac{2,5}{3} \sigma) \approx 0,4$$

SAMPLING FROM CONTAMINATED DISTRIBUTIONS

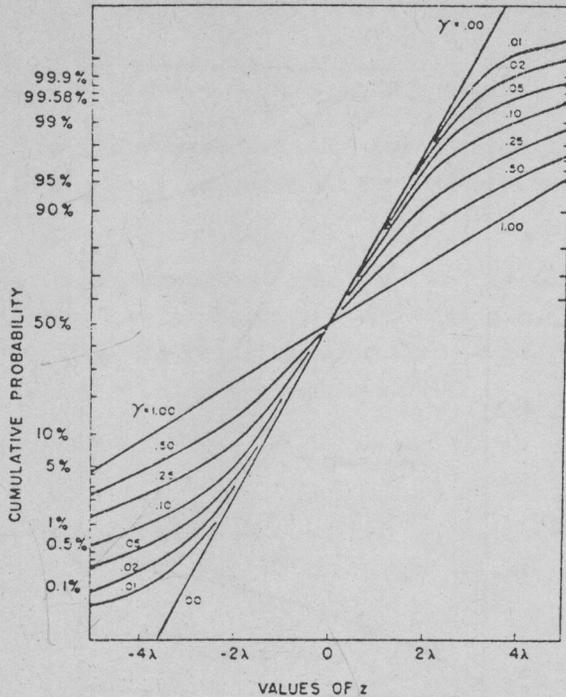


Fig.2. Cumulatives of Normal and Contaminated Distributions

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. Donc à moins qu'une ou plusieurs de ces quatre valeurs ne soit très grande, il ne sera guère possible de détecter que le modèle normal ne convient pas. Or l'efficacité de \bar{X}_n décroît assez vite lorsque la fraction de contamination ϵ croît de 0 à 0,10 :

ϵ				
Fraction de contamination	0,00	0,02	0,05	0,10
Efficacité de \bar{X}_n	1,00	0,90	0,80	0,70

Par contre, si l'on ôte une fraction 2α des observations, les $[\alpha n]$ plus grandes et les $[\alpha n]$ plus petites ($[u]$ désignant la partie entière de u), et si l'on prend la moyenne des observations restantes, obtenant ainsi la moyenne α -tronquée des observations, on obtient une

plus grande stabilité de l'efficacité asymptotique.

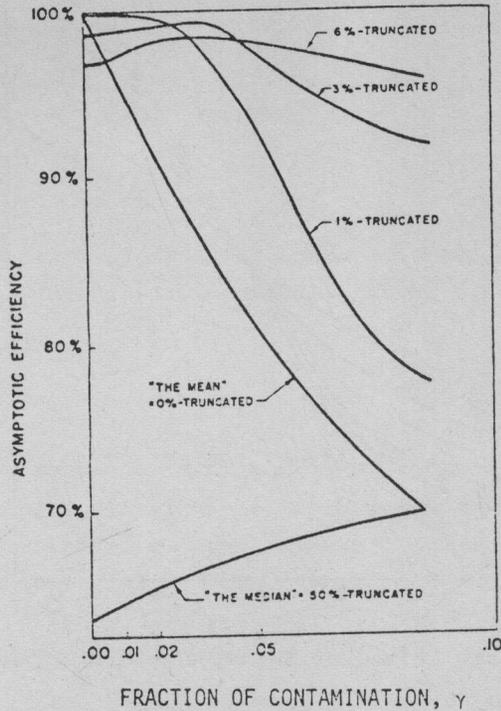


Fig.4. Asymptotic Efficiency, for Location, of Truncated Mean

En tronquant à 1%, la perte d'efficacité pour le modèle strictement normal est minime (0,4%). Par contre, lorsque ϵ varie entre 0 et 0,10, le gain peut dépasser 12%.

Tukey a même obtenu le résultat suivant : 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 $\epsilon E [0 ; 0,10]$. L'un quelconque de ces estimateurs est donc préférable à \bar{X} qui voit son efficacité décroître de 1 à 0,7 dans cette même zone.

2 - ESTIMATION D'UN PARAMETRE D'ECHELLE ET CONTAMINATION

Etant donné un n-échantillon (X_1, \dots, X_n) d'une variable X de loi $F\left(\frac{\cdot - \theta}{\sigma}\right)$, où θ et σ sont inconnus, on considère les deux estimateurs sans biais de σ fondés sur les mesures suivantes de la dispersion de l'échantillon :

$$d_n = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

$$s_n = \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}$$

Dès 1914, comme le remarque P.J. Huber (1981), il y a eu une controverse entre Eddington, qui arguait en faveur du premier et Fisher en faveur du second :

En fait, si le second est optimal dans le cas où F est normale, le premier résiste mieux à la présence d'une contamination, comme le montre un calcul de l'efficacité relative asymptotique de D_n par rapport à S_n pour le modèle contaminé précédent :

$$F_\varepsilon = (1-\varepsilon)N(0, \sigma^2) + \varepsilon N(0, 9\sigma^2)$$

$$\begin{aligned} \text{ERA}_{D_n/S_n}(\varepsilon) &= \lim_{n \rightarrow \infty} \frac{\text{Var}(S_n) / (ES_n)^2}{\text{Var}(D_n) / (ED_n)^2} \\ &= \frac{\left(\frac{3(1+80\varepsilon)}{(1+8\varepsilon)^2} - 1 \right) / 4}{\frac{\pi(1+8\varepsilon)^2}{2(1+2\varepsilon)^2} - 1} \end{aligned}$$

ε	0	0,001	0,002	0,01	0,05	0,50
Fraction de contamination	0	0,001	0,002	0,01	0,05	0,50
Efficacité relative asymptotique de D_n/S_n	0,876	0,948	1,016	1,016	1,439	1,017

Extrait de la table de P.J. Huber (1977)

Il suffit donc d'une contamination de 2‰ pour que l'écart moyen absolu soit meilleur que l'écart moyen quadratique.

3 - EXEMPLES D'ANALYSE DE VARIANCE

Nous allons voir sur trois exemples (Scheffee (1959) les effets de la violation des hypothèses suivantes :

- Normalité des erreurs
- Egalité des variances
- Indépendance des erreurs

a) Non normalité

La non normalité d'une loi sera mesurée à partir des moments centrés d'ordre trois et quatre de la variable correspondante, le premier donnant une indication sur l'asymétrie de la loi et le second sur son aplatissement, par les deux quantités suivantes invariantes par translation et par changement d'échelle :

$$\gamma_1 = \frac{1}{\sigma^3} E[(X - EX)^3] \quad \text{Coefficient d'asymétrie (skewness en anglais)}$$

$$\gamma_2 = \frac{1}{\sigma^4} E[(X - EX)^4] - 3 \quad \text{Coefficient d'aplatissement (kurtosis en anglais)}$$

De ces deux quantités, nulles pour une loi gaussienne, la première est positive si la loi est déportée sur la droite, la seconde est positive si la loi a des queues importantes.

Dans le cas gaussien, pour tester l'hypothèse $\sigma^2 = \sigma_0^2$, on utilise la statistique

$$Q = \frac{\sum (X_i - \bar{X})^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2}$$

qui sous l'hypothèse nulle ($\sigma^2 = \sigma_0^2$), suit une loi du χ^2 à $n-1$ degrés de liberté, d'espérance $n-1$ et de variance $2(n-1)$, pourvu que la loi de X soit gaussienne. Sinon

$$E\left(\frac{S^2}{\sigma_0^2}\right) = 1 \quad \text{Var}\left(\frac{S^2}{\sigma_0^2}\right) = \frac{2}{n-1} + \frac{\gamma_2}{n}$$

On doit donc fonder le test sur la statistique $\left(\frac{n-1}{2}\right)^{1/2} \left(\frac{S^2}{\sigma_0^2} - 1\right)$ qui, sous l'hypothèse nulle est asymptotiquement gaussienne $N(0, 1 + \frac{1}{2}\gamma_2)$. Tests et intervalles de confiance sont donc modifiés. Le tableau suivant (Scheffé p.337) donne, pour l'intervalle de confiance de 95% obtenu en considérant que la loi de X est gaussienne, la probabilité α qu'il ne contienne pas σ^2 , pour plusieurs valeurs de l'aplatissement γ_2 .

γ_2	-1,5	-1	-0,5	0	0,5	1	2	4	7
α	$9 \cdot 10^{-5}$	0,006	0,024	0,050	0,080	0,11	0,17	0,26	0,36

b) Dépendance

L'observation (X_1, \dots, X_n) est un vecteur gaussien, dont toutes les composantes X_i ont la même loi $N(\theta, \sigma^2)$ et dont les seules cova-

riances non nulles sont celles de deux termes consécutifs :

$$\text{cov}(X_i, X_{i+1}) = \rho\sigma^2 \quad i = 1, 2, \dots, n-1.$$

Les tests de l'hypothèse $(\theta = \theta_0)$, fondés sur la statistique

$$T = \frac{(\bar{X} - \theta_0)\sqrt{n}}{S}, \text{ dont la loi est celle de Student à } n-1 \text{ degrés de liberté}$$

sous l'hypothèse nulle lorsque $\rho = 0$, et donc asymptotiquement $N(0,1)$, vont être affectés par cette dépendance car :

$$\bar{E}\bar{X} = \theta, \text{ mais } \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \left[1 + 2\rho\left(1 - \frac{1}{n}\right)\right] \text{ et}$$

$$E(S^2) = \sigma^2\left(1 - \frac{2\rho}{n}\right).$$

Par suite T est asymptotiquement gaussienne $N(0, 1 + 2\rho)$ au lieu de $N(0,1)$.

Le tableau ci-dessous est l'analogie du précédent en fonction du coefficient de corrélation ρ , pour l'intervalle de confiance à 95% de la moyenne :

ρ	-0,4	-0,3	-0,2	-0,1	0	0,1	0,2	0,3	0,4
α	10^{-5}	0,002	0,011	0,028	0,050	0,074	0,098	0,12	0,14

c) Variances inégales

Le problème est celui de la comparaison de deux échantillons indépendants de tailles respectives n_1 et n_2 , X_1, \dots, X_{n_1} et Y_1, \dots, Y_{n_2} , des variables X et Y dont les lois ont pour moyennes respectives θ_1 et θ_2 et pour variances respectives σ_1^2 et σ_2^2 .

On note r le rapport des tailles des deux échantillons, q celui des variances et Δ l'écart entre les moyennes :

$$r = \frac{n_1}{n_2} \quad q = \frac{\sigma_1^2}{\sigma_2^2} \quad \Delta = \theta_1 - \theta_2$$

Les tests de l'hypothèse H_0 ($\Delta = \Delta_0$) sont fondés sur la statistique

$$T = \frac{\bar{X} - \bar{Y} - \Delta_0}{\left(\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}\right)^{1/2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}}$$

Sous l'hypothèse nulle et lorsque n_1 et n_2 sont grands le numé-

teur de T est approximativement normal $N(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$. Mais le dénominateur, où S_1^2 et S_2^2 convergent en probabilité vers σ_1^2 et σ_2^2 , ne donne pas la normalisation convenable pour que T soit asymptotiquement normal $N(0,1)$:

T est asymptotiquement gaussien $N(0, \frac{r+q}{rq+1})$.

Donc si les tailles des deux échantillons sont égales, l'inégalité des variances n'a aucun effet sur le comportement asymptotique de T .

Le tableau ci-dessous donne la probabilité α pour que l'intervalle de confiance à 95% construit pour $\Delta = \theta_1 - \theta_2$ sous l'hypothèse de normalité et d'égalité des variances ne contienne pas Δ en fonction du rapport q des variances et de celui r des tailles des deux échantillons.

r \ q	0	1/5	1/2	1	2	5	∞
1	0,05	0,05	0,05	0,05	0,05	0,05	0,05
2	0,17	0,12	0,08	0,05	0,03	0,01	0,00
5	0,38	0,22	0,12	0,05	0,01	0,00	0,00
∞	1,00	0,38	0,17	0,05	0,00	0,00	0,00

4 - COMPARAISON DES TESTS DE WILCOXON ET DE STUDENT

Le problème est à nouveau celui de la comparaison de deux échantillons indépendants

X_1, \dots, X_{n_1} d'une variable X de fonction de répartition F_1

Y_1, \dots, Y_{n_2} Y F_2

On veut tester $H_0 (F_2 = F_1)$ contre l'alternative de translation ($F_2 = F_1(\cdot - \theta)$, $\theta > 0$).

Si l'on sait que F_1 est une loi gaussienne $N(\theta_0, \sigma^2)$, θ_0 et σ^2 inconnus, le test de Student fondé sur T est uniformément le plus puissant :

$$T = \frac{(\bar{Y} - \bar{X})(n_1 + n_2 - 2)^{1/2}}{(\sum (Y_i - \bar{Y})^2 + \sum (X_i - \bar{X})^2)^{1/2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{1/2}}$$

Mais si l'on ne connaît pas F_1 , on peut employer le test de Wilcoxon,

l'un des plus simples parmi les tests de rangs de H_0 ($F_2 = F_1$) :
 Il consiste à rejeter H_0 si la somme W des rangs des Y_i dans la suite $(X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2})$ ordonnée par ordre croissant est trop grande.

$$W = \sum_{i=1}^{n_2} \left(\sum_{j=1}^{n_2} 1_{Y_i \geq Y_j} + \sum_{j=1}^{n_1} 1_{Y_i \geq X_j} \right)$$

Ces tests sont à niveau constant pour toutes les lois de fonction de répartition F_1 continue et sont localement les plus puissants parmi les tests de rangs pour la loi logistique (Hájek 1967). On note $n = n_1 + n_2$.

Pour comparer ces deux tests, on calcule leur efficacité relative asymptotique (ERA), définie comme la limite de l'inverse du rapport des vitesses de convergence de l'alternative vers l'hypothèse nulle ($\theta_T(n) \rightarrow 0, \theta_W(n) \rightarrow 0$) qui maintiennent constants le niveau et la puissance pour l'un et l'autre test. Les deux statistiques T et W sont asymptotiquement gaussiennes et la convergence est suffisamment régulière pour que leur efficacité relative asymptotique au point F_1 puisse se calculer comme (Cramer (1974), Raoult (1976)) :

$$e(W/T) = \lim_{n_1 n_2 \rightarrow \infty} \left[\frac{\frac{\partial}{\partial \theta} E_{\theta} W_n}{\sigma_{\theta}(W_n)} \cdot \frac{\sigma_{\theta}(T_n)}{\frac{\partial}{\partial \theta} E_{\theta}(T_n)} \right]_{\theta=0}^2$$

Supposons que F_1 soit la fonction de répartition d'une loi admettant une densité f_1 par rapport à la mesure de Lebesgue sur \mathbb{R} de carré intégrable, et un moment d'ordre deux, égal à σ^2 ($E_{F_1} X$ supposé nul). n_1 et n_2 tendent vers l'infini de telle sorte que $n_1 = \lambda n + o(n)$ et $n_2 = (1-\lambda)n + o(n)$ pour un λ de $]0, 1[$; $p(\theta) = P(Y > X) = \int F_1(x-\theta) f_1(x) dx$.

$$\begin{aligned} \text{Alors } E_{\theta}(\bar{Y}_{n_2} - \bar{X}_{n_1}) &= \theta \\ \sigma_{\theta}^2(\bar{Y}_{n_2} - \bar{X}_{n_1}) &= \frac{\sigma^2}{n} \left(\frac{1}{1-\lambda} + \frac{1}{\lambda} \right) + \theta \left(\frac{1}{n} \right) \\ E_{\theta}(W_{n_1, n_2}) &= \lambda(1-\lambda)n^2 p(\theta) + \frac{(1-\lambda)^2 n^2}{2} \\ \sigma_{\theta}^2(W_{n_1, n_2}) &= \frac{1}{12} n^2 \lambda(1-\lambda)(n+1) \end{aligned}$$

Finalement

$$e_{F_1}(W/T) = 12 \sigma^2 \left(\int f_1^2(x) dx \right)^2$$