ANNALES SCIENTIFIQUES
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Mathématiques

L. BIRGE

Tests robustes pour des variables indépendantes et des chaînes de Markov

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 71, série *Mathématiques*, n° 20 (1982), p. 70-77

<http://www.numdam.org/item?id=ASCFM_1982__71_20_70_0>

TESTS ROBUSTES POUR DES VARIABLES INDEPENDANTES ET DES CHAINES DE MARKOV

L. BIRGE
Université de PARIS X-NANTERRE

I - INTRODUCTION

Le problème de l'obtention de tests robustes a été abordé par de nombreux auteurs sous divers aspects. Celui qui nous intéressera ici est le suivant : étant donnés deux voisinages $\mathcal{P}$ et $\mathcal{Q}$ de P et Q respectivement, peut-on obtenir de "bons" tests, voire des tests minimax de $\mathcal{P}^{\otimes n}$ contre $\mathcal{Q}^{\otimes n}$, où $\mathcal{P}^{\otimes n}$ désigne le produit de n voisinages identiques à $\mathcal{P}$. Ce problème a été largement étudié dans deux articles par Huber et Strassen [6] [7] lorsque P et Q sont des ensembles convexes engendrés par des capacités bi-alternatives. Ces travaux ont été généralisés par Birgé [2], Rieder [11] et Bednarski [1]. Les résultats montrent l'existence d'un couple $(\overline{P}, \overline{Q})$ le moins favorable et les tests de rapport de vraisemblance de ce couple sont minimax de $\mathcal{P}^{\otimes n}$ contre $\mathcal{Q}^{\otimes n}$ quel que soit le niveau et le nombre n des observations.

De tels résultats ne subsistent plus lorsque P et Q ne sont pas engendrés par des capacités bi-alternatives. En particulier ils sont faux pour les boules en distance de Hellinger définie par :

$$h^2(P,Q) = \frac{1}{2}\int(\sqrt{dP} - \sqrt{dQ})^2 .$$

Toutefois, pour des convexes compacts on peut encore trouver de bons tests, dont les erreurs décroissent en fonction exponentielle du nombre des observations, les coefficients de ces exponentielles étant asymptotiquement optimaux. Ces résultats sont démontrés dans Birgé [3]. En particulier dans le cas des boules en distance de Hellinger, les tests prennent une forme explicite particulièrement simple.

Cependant tous ces résultats reposent de façon essentielle sur la convexité des ensembles $\mathcal{P}$ et $\mathcal{Q}$ considérés. Or le fait d'utiliser des produits tels que $\mathcal{P}^{\otimes n}$ revient à dire que l'on observe n variables indépendantes $X_1, \ldots, X_n$ de lois respectives $P_1, \ldots, P_n$ et que chacune de ces lois est assez près de P pour appartenir à $\mathcal{P}$. Ceci exclut en fait le cas des "outliers". Si l'on considère le problème sous un autre aspect, les voisinages obtenus seront très différents. Supposons qu'effectivement la plupart des lois P_i sont dans un voisinage de P mais qu'un petit nombre d'entre elles en sont très éloignées, sans que l'on sache lesquelles (ce qui permettrait d'éliminer les observations correspondantes). Une telle conception de la robustesse nous amène à modéliser les voisinages de façon différente et de ne plus considérer des produits de voisinages, mais directement des voisinages dans l'espace des mesures produits en utilisant une distance du type de celle introduite par Le Cam (d'ailleurs dans un contexte un peu différent) et définie ainsi :

$$H^2(P_1 \otimes \ldots \otimes P_n, Q_1 \otimes \ldots \otimes Q_n) = \sum_{i=1}^{n} h^2(P_i, Q_i) .$$

Si l'on note $\mathcal{B}_d(P, r)$ la boule fermée de centre P et rayon r pour la distance d, il est immédiat que l'on a :

$$\mathcal{B}_H(P^{\otimes n}, r) \subset \mathcal{B}_h^{\otimes n}(P, \frac{r}{\sqrt{n}})$$

mais que l'inclusion est stricte et que la boule en distance H ne peut s'écrire ni comme un convexe ni comme un produit de convexes. Il s'ensuit qu'aucune des techniques de minimax utilisées habituellement pour ce type de problème ne peut s'appliquer. Les difficultés sont autres lorsqu'il s'agit de définir des tests robustes entre des chaînes de Markov car c'est alors la définition des voisinages qui pose des problèmes.

Pourtant dans ces deux cas, il est désirable d'obtenir, à défaut de tests minimax des tests robustes dont les erreurs décroissent exponentiellement vite en fonction du nombre des observations. Outre leur intérêt propre, l'existence de tels tests est essentielle pour la construction d'estimateurs du type envisagé par Le Cam [8] , [9] et Birgé [4] . Le Cam a construit des tests entre des boules en distance H ayant les propriétés voulues mais il est obligé de se restreindre à des rayons inférieurs à 0,05, ce qui ne permet pas de donner aux estimateurs les performances recherchées. La méthode ne peut malheureusement pas se généraliser et il faut aborder le problème d'une toute autre façon. Les résultats qui suivent concernent donc la construction de tests dans le cas de la distance H pour des variables indépendantes ainsi que dans le cas de chaînes de Markov, avec quelques hypothèses supplémentaires, en utilisant une distance définie ci-dessous.

II - RAPPELS ET PRÉLIMINAIRES TECHNIQUES :

Soient $(\Omega, \mathcal{A})$ un espace métrique localement compact muni de sa tribu borélienne et $\mathcal{M}_1$ l'ensemble des probalités sur Ω, muni de la distance de Hellinger h. Considérons les boules $\mathcal{B}(P, \varepsilon)$ et $\mathcal{B}(Q, \eta)$ ε, η non tous deux nuls avec $h(P, Q) > \varepsilon + \eta$. Il a été montré dans [3] que si l'on définit $\overline{P}$ et $\overline{Q}$ de la façon suivante

$$\sqrt{\frac{d\overline{P}}{d\mu}} = \alpha\sqrt{\frac{dP}{d\mu}} + \alpha'\sqrt{\frac{dQ}{d\mu}} \quad ; \quad \sqrt{\frac{d\overline{Q}}{d\mu}} = \beta'\sqrt{\frac{dP}{d\mu}} + \beta\sqrt{\frac{dQ}{d\mu}}$$

avec

$$\mu = P + Q \quad ; \alpha^2 + \alpha'^2 + 2\,\alpha\alpha'(1-a) = 1 \quad ; \beta^2 + \beta'^2 + 2\,\beta\beta'(1-a) = 1 \quad ;$$
$$\alpha + \alpha'(1-a) = 1-\varepsilon^2 \quad ; \beta + \beta'(1-a) = 1-\eta^2 \quad ; a = h^2(P,Q) \quad ,$$

on peut choisir une version Φ du rapport de vraisemblance $\frac{d\overline{Q}}{d\overline{P}}$ ayant la propriété suivante :

<u>Théorème</u> : Quel que soit P' dans $\mathcal{B}(P, \varepsilon)$ et Q' dans $\mathcal{B}(Q, \eta)$

$$\int \sqrt{\Phi}\; dP' \leq \int \sqrt{\Phi}\, d\overline{P} \quad ; \quad \int \Phi^{-\frac{1}{2}}\, dQ' \leq \int \Phi^{-\frac{1}{2}}\, d\overline{Q} \;.$$

Il s'ensuit aisément en utilisant une inégalité exponentielle que l'on a :

<u>Corollaire</u> : Les tests de rapports de vraisemblance au seuil 0 de $\overline{P}^{\otimes n}$ contre $\overline{Q}^{\otimes n}$ sont des tests de $\mathcal{B}^{\otimes n}(P, \varepsilon)$ contre $\mathcal{B}^{\otimes n}(Q, \eta)$ dont les erreurs sont toutes deux majorées par α avec :

$$(0) \qquad \alpha \leq \exp\{-nh^2(\overline{P},\overline{Q})\} \quad ; \quad h(\overline{P},\overline{Q}) > h(P,Q) - \varepsilon - \eta \;.$$

On obtient ainsi un "bon" test entre les deux boules, mais dans la mesure où le choix du rayon ε est plus ou moins arbitraire on peut se demander ce que deviennent les performances d'un tel test si certaines des observations ont des lois dont la distance à P (ou à Q) est supérieure à ε. Dans le cas où $\eta = \varepsilon$, le lemme suivant permet d'obtenir la solution de ce problème et constitue le résultat fondamental pour la suite de cette étude. Son expression sera plus aisée si l'on utilise un paramétrage particulier. Nous poserons donc :

$$h^2(P,Q) = 2\sin^2\left(\frac{\alpha}{2}\right) \quad \left(0 < \alpha \leq \frac{\pi}{2}\right) \quad ; \quad \varepsilon^2 = h^2(P,\bar{P}) = 2\sin^2\left(\frac{\lambda\alpha}{2}\right) ,$$

et nous supposerons que $0 < \lambda < \frac{1}{2}$ ce qui implique que les boules en question sont disjointes. Il s'ensuit aisément que $\bar{P}$ et $\bar{Q}$ seront définis par

$$\sin\alpha\sqrt{\frac{d\bar{P}}{d}} = \sin\lambda\alpha\sqrt{\frac{dQ}{d}} + \sin(1-\lambda)\alpha\sqrt{\frac{dP}{d}}$$

$$\mu = P + Q$$

$$\sin\alpha\sqrt{\frac{d\bar{Q}}{d}} = \sin\lambda\alpha\sqrt{\frac{dP}{d}} + \sin(1-\lambda)\alpha\sqrt{\frac{dQ}{P}}$$

Soit $\Phi = \dfrac{d\bar{Q}}{d\bar{P}}$ comme précédemment, alors pour toute probabilité P' on a le résultat suivant :

<u>Lemme</u> : Définissons les quantités $A \geq \frac{2}{1-\lambda}$, $K(\lambda) = \frac{2A}{A-2} \times \frac{1-2\lambda}{\lambda}$,

$$B(\lambda) = \frac{\sin^2\frac{\pi}{4}(1-2\lambda)}{(1-2\lambda)^2} \quad \text{alors}$$

i) si $h(P,P') \leq h(P,\bar{P}) = \sqrt{2}\,\sin\frac{\lambda\alpha}{2}$, en particulier si $h(P,P') \leq \lambda h(P,Q)$ on a

(1) $$\int\sqrt{\Phi}\,dP' \leq 1 - (1-2\lambda)^2\,h^2(P,Q).$$

ii) si $h(P,P') > h(P,\bar{P})$ on a

$$\int\sqrt{\Phi}\,dP' \leq 1 + K(\lambda)h^2(P,P') - (1 - \lambda A(A-2)\,B(\lambda))\,(1-2\lambda)\,h^2(P,Q).$$

Il est facile de voir que si $A = 1+\sqrt{5}$ (ce qui semble un assez bon choix pour des valeurs de λ entre 0.1 et 0.3) cette dernière équation peut s'écrire

$$\int\sqrt{\Phi}\,dP' \leq 1 + (3+\sqrt{5})\,\frac{1-2\lambda}{\lambda}\,h^2(P,P') - (1-2\lambda)\left[1 - \frac{4\lambda\sin^2\left[\frac{\pi}{4}(1-2\lambda)\right]}{(1-2\lambda)^2}\right]h^2(P,Q).$$

Il est clair qu'une telle inégalité devient sans intérêt pour de très petites valeurs de λ, ce qui n'est guère surprenant car alors le rapport de vraisemblance $\dfrac{d\bar{Q}}{d\bar{P}}$ peut prendre des valeurs très grandes, mais qu'elle donne des résultats satisfaisants pour de moyennes valeurs de λ.

Si $X_1, \ldots, X_n$ sont tirés selon des lois $P_1, \ldots, P_n$, il est aisé de voir que le niveau du test obtenu en utilisant comme statistique de test le rapport de vraisemblance de $\bar{Q}^n$ et $\bar{P}^n$ sera inférieur à $\prod_{i=1}^{n} \int \sqrt{\Phi}\, dP_i$. Donc si λ est de l'ordre de 0.2 et si un petit nombre seulement des P_i sont légèrement à l'extérieur de la boule $\mathcal{B}(P, \varepsilon)$, les performances du test ne seront pas trop perturbées comme on peut le voir grâce aux inégalités (1) et (2). C'est cet approche empirique que nous allons rendre plus rigoureux en utilisant la distance H.

III - CAS DES VARIABLES INDÉPENDANTES NON ÉQUIDISTRIBUÉES :

Soient deux mesures produits $P = P_1 \otimes \ldots \otimes P_n$ et $Q = Q_1 \otimes \ldots \otimes Q_n$, $\mathcal{P} = \mathcal{B}_H(P, r)$ et $\mathcal{Q} = \mathcal{B}_H(Q, r)$ deux boules entre lesquelles on désire tester. Fixons $0 < \lambda \leq 0.37$. Pour $1 \leq i \leq n$, on peut construire $\bar{P}_i$ et $\bar{Q}_i$ à partir de P_i, Q_i et λ comme il a été indiqué précédemment. Soit $\Phi_i = \frac{d\bar{Q}_i}{d\bar{P}_i}$,

on utilisera pour tester $\mathcal{P}$ contre $\mathcal{Q}$ la statistique $\Psi = \sum_{i=1}^{n} \log \Phi_i(x_i)$ en rejetant $\mathcal{P}$ lorsque Ψ est positive. En utilisant le lemme ci-dessus on peut démontrer le théorème suivant.

<u>Théorème</u> : Soit Ψ défini comme il a été indiqué ; pour tout P' dans $\mathcal{P}$ et Q' dans $\mathcal{Q}$ on a :

$$P'\left[\Psi \geq 0\right] \leq \alpha \qquad \text{et} \qquad Q'\left[\Psi \leq 0\right] \leq \alpha \quad , \qquad \text{avec}$$

$$\alpha = \exp\left[-(1-2\lambda)^2 H^2(P,Q) + \frac{1-2\lambda}{\lambda} B(\lambda) \left[\bar{A}(\lambda)-2\right]\left[3\bar{A}(\lambda) - 2\right] r^2\right]$$

où $\bar{A}(\lambda)$ est la solution positive de l'équation $(x-2)^2(x-1)\, B(\lambda) = 2$.

Il suffit alors de choisir λ convenablement pour obtenir le corollaire suivant :

<u>Corollaire</u> : Supposons que $H(P, Q) \geq C\, r$. Pour tout $C \geq 7$, il existe une constante $K(C)$ telle que l'on ait pour un choix judicieux de λ

$$\alpha \leq \exp\left[-K(c)\, H^2(P,Q)\right] \quad ,$$

avec en particulier si $C = 7$ $K(c) > \frac{1}{40}$ et si $C = 16,52$ $K(c) > \frac{1}{2}$.

On peut comparer ce résultat à celui obtenu avec la distance de Hellinger. Lorsque les variables sont équidistibuées, $P = P_1^{\otimes n}$, $Q = Q_1^{\otimes n}$ de sorte que $H(P, Q) = \sqrt{n}\, h(P_1, Q_1)$. On retrouvera une inégalité absolument analogue à (0) sauf que la constante est un peu différente.

IV - LE PROBLÈME DES CHAÎNES DE MARKOV :

Supposons que nous disposions d'un ensemble $\mathfrak{M}$ de noyaux de transitions $P(x,.) = P_x^1$; à chaque noyau nous pouvons associer la suite de ses itérés P_x^n, que nous désignerons par P. Etant donnée une mesure initiale, on peut faire correspondre à P une chaîne de Markov. On sait très bien tester entre deux telles chaînes par le lemme de Neyman-Pearson, mais si l'on veut parler de robustesse il faut définir des voisinages de P. Or ce problème n'est pas du tout immédiat car il n'existe pas de distance "naturelle" entre deux noyaux de transition. On peut bien sûr considérer $h(P_x^1 , Q_x^1)$ mais ceci dépend de x (il en est de même si on utilise une autre distance que h) et il est clair que prendre l'inf ou le sup en x ne donnera pas une distance correspondant à une quelconque notion de séparation entre les deux chaînes, alors que la distance obtenue doit être liée aux erreurs des tests de P contre Q. Pour la construire, nous ferons les hypothèses suivantes sur $\mathfrak{M}$.

<u>H1</u> : Il existe un entier k et une mesure positive μ tels que pour tout x et tout P

$$\mu(.) \leqslant k^{-1} \sum_{i=1}^{k} P_x^k(.) \quad .$$

<u>H2</u> : Il existe un entier l et une mesure positive finie ν tels que pour tout x et tout P

$$P_x^1(.) \leqslant \nu(.) \quad .$$

Définissons alors pour toute mesure positive λ

$d_\lambda^2(P,Q) = \int h^2(P_x^1 , Q_x^1)\, d\lambda(x)$. Considérons deux chaînes P et Q avec $d_\mu(P, Q) = R$ et deux boules $\mathcal{P} = \mathcal{B}(P, r)$, $\mathcal{Q} = \mathcal{B}(Q, r)$ pour la distance d_ν.

Théorème : Soient $R \geqslant Cr$ avec $C \geqslant 7$ et $n \geqslant k+1$. Il existe un test de $\mathcal{P}$ contre $\mathcal{Q}$ fondé sur $X_0, \ldots, X_n$ dont les erreurs sont majorées par

$$\exp\left[-\left[\!\left[\frac{n}{k+1}\right]\!\right] K(C)\, R^2\right] \qquad \text{où } [\![u]\!] \text{ désigne la partie entière de } u.$$

On obtient donc une généralisation des résultats précédents aux chaînes de Markov.

V - CONCLUSION :

On voit ainsi que dans tous ces cas, la vitesse de séparation de deux hypothèses constituées par des boules, en fonction du nombre des observations, prend la forme $\exp\left[-\,And^2\right]$ où d est une distance convenable. Des résultats semblables ont été également donnés dans le cas de processus gaussiens stationnaires (cf. [4]) mais sous des hypothèses trop fortes. Outre leur intérêt propre en tant que moyen d'obtenir des tests robustes de construction explicite, les résultats précédents s'appliquent à l'estimation dans des constructions telles que celle de Le Cam [9] et Birgé [4]. Pour que ces constructions soient possibles, l'hypothèse fondamentale est justement une hypothèse de séparation des boules du type précédent, ce qui fait que les théorèmes ci-dessus permettent d'étendre les résultats obtenus pour les variables indépendantes équidistribuées aux cas des variables non équidistribuées et des chaînes de Markov.

BIBLIOGRAPHIE

1) BEDNARSKI T. (1978) : Binary experiments. Minimax tests and two-alternating capacities. Institute of Mathematics, Polish Ac. of Sc. - Preprint 131.

2) BIRGÉ L. (1977) : Tests minimax robustes, Séminaire de Statistiques d'Orsay 1974-1975, chap. VII. Astérisque 43-44.

3) BIRGÉ L. (1979) : Sur un théorème de minimax et son application aux tests. A paraître dans Probability and Mathematical Statistics.

4) BIRGÉ L. (1980) : Approximation dans les espaces métriques et théorie de l'estimation (à paraître).

5) CHERNOFF H. (1952) : A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statistics 23, 493-507.

6) HUBER P.J. (1965) :A robust version of the probability ratio test. Ann. Math. Statistics 36, 1753-1758.

7) HUBER P.J. et STRASSEN V. (1973) : Minimax tests and the Neymann-Pearson lemma for capacities. Ann. of Statistics 1, 251-263.

8) LE CAM L. (1973) : Convergence of estimates under dimensionality restrictions. Ann. of Statistics 1, 38-53.

9) LE CAM L. (1975) : On local and global properties in the theory of asymptotic normality of experiments. Stochastic Processes and Related Topics 1, 13-54. Academic Press - New-York.

10) LEHMANN E.L. (1959) : Testing Statistical Hypotheses. J. Wiley - New-York.

11) RIEDER H. (1977) : Least favorable pairs for special capacities. Ann. of Statistics 5, 909-921.

Lucien BIRGÉ
Université Paris X - Nanterre,
U.E.R. d'Economie,
200 Avenue de la République
92001 - Nanterre CEDEX.
E.R.A. C.N.R.S. 532 (Orsay).