

PRINCIPE DE SYMÉTRISATION POUR LA CONSTRUCTION D'UN TEST ADAPTATIF

Cécile Durot¹ & Yves Rozenholc²

¹ *UFR SEGMI, Université Paris Ouest Nanterre La Défense, France,
cecile.durot@gmail.com*

² *Université Paris Descartes, MAP5 - UMR CNRS 8145 et INRIA Saclay Ile-de-France
équipe Select, France, yves.rozenholc@parisdescartes.fr*

Résumé. Ayant observé un vecteur aléatoire y de \mathbb{R}^n , et notant f son espérance, nous bâtissons un test non-asymptotique de l'hypothèse nulle que $f = 0$ contre l'alternative $f \neq 0$, sous la seule hypothèse que les composantes y_i de y sont indépendantes entre elles et de loi symétrique autour de leur espérance. Nous ne supposons pas les composantes $y_i - f_i$ de même loi. Sans hypothèse supplémentaire, la loi d'une statistique de test est inconnue même sous l'hypothèse nulle. Nous proposons donc une calibration fondée sur un principe de symétrisation exploitant l'hypothèse de symétrie. Le test est non-asymptotique. Sa puissance est décrite dans le cadre du modèle de régression, où $f_i = F(x_i)$ pour une fonction F inconnue et des $x_i \in [0, 1]$ déterministes : le test est adaptatif sur des classes de régularité Hölderienne sous des hypothèses générales de moments.

Cette présentation est effectuée pour la session spéciale du groupe statistique mathématique sur les techniques de symétrisation.

Mots-clés. test adaptatif, test minimax, alternative nonparamétrique, symétrisation, hétéroscédasticité.

Abstract. We observe a random vector y in \mathbb{R}^n . With $f = \mathbb{E}(y)$, we build a non-asymptotic test for the null hypothesis $f = 0$ against the alternative hypothesis that $f \neq 0$ when the components y_i of y are independent with a distribution that is symmetrical about its expectation. The components $y_i - f_i$ are not assumed to be identically distributed. The distribution of a given test statistic is unknown even under the null hypothesis due to the lack of assumptions on the distribution of y . Hence, we propose a calibration that is based on a symmetrization principle that exploits the symmetry assumption. The test is non-asymptotic. Its power is described in the regression setting, where $f_i = F(x_i)$ for an unknown regression function F and fixed design points $x_i \in [0, 1]$. The test is adaptive with respect to Hölderian smoothness in the regression setting under mild assumptions on ε .

This presentation is for the special session on symmetrization principles.

Keywords. adaptive test, minimax hypothesis testing, nonparametric alternatives, symmetrization, heteroscedasticity.

1 Construction du test

Nous observons un vecteur aléatoire $y \in \mathbb{R}^n$ et notons $y = f + \varepsilon$, où f est l'espérance inconnue de y et $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ est un vecteur aléatoire centré non observé. Nous supposons que les ε_i , $i = 1, \dots, n$, sont mutuellement indépendantes et de loi symétrique, c'est-à-dire que ε_i a la même loi que $-\varepsilon_i$. En revanche, nous ne supposons pas que les ε_i sont de même loi. Notre objectif est de bâtir un test non-asymptotique de

$$H_0 : f = 0 \quad \text{contre} \quad H_1 : f \neq 0$$

sans autre hypothèse sur f ou ε . Nous proposons un multi-test dont la calibration repose sur un principe de symétrisation.

1.1 Région de rejet

Construisons tout d'abord une forme de région de rejet. Pour ce faire, nous considérons une collection \mathcal{M} de partitions m de $\{1, \dots, n\}$ en D_m parties non vides. Pour tout $m \in \mathcal{M}$, nous notons Π_m le projecteur orthogonal de \mathbb{R}^n sur l'espace vectoriel constitué des vecteurs u dont les composantes sont constantes par morceaux sur chacune des D_m parties (pour tout $e \in m$ et tous $i, j \in e$, $u_i = u_j$). Notons que la dimension de cet espace est D_m et que pour tout u dans cet espace,

$$\|\Pi_m u\|_n^2 = \sum_{e \in m} \frac{1}{|e|} \left(\sum_{k \in e} u_k \right)^2$$

où $\|\cdot\|_n$ désigne la norme euclidienne dans \mathbb{R}^n et $|e|$ désigne le cardinal de e .

La construction de la région de rejet repose sur le fait que sous H_0 , $\|\Pi_m y\|_n^2 = \|\Pi_m \varepsilon\|_n^2$ a tendance à être petit pour tout m tandis que sous H_1 , nous pouvons typiquement trouver (au moins) une partie $e \subset \{1, \dots, n\}$ telle que $(\sum_{k \in e} y_k)^2$ ait tendance à être grande, de sorte que $\|\Pi_m y\|_n^2$ ait elle aussi tendance à être grande si $e \in m$. Le test consiste alors à rejeter H_0 s'il existe au moins une partition m dans la collection \mathcal{M} telle que $\|\Pi_m y\|_n^2$ soit trop grande. Cela signifie que pour chaque $m \in \mathcal{M}$, nous réalisons un test de niveau α_m consistant à rejeter H_0 si $\|\Pi_m y\|_n^2$ exède un certain seuil $q_m^y(\alpha_m)$, choisi de telle sorte que

$$\mathbb{P}_{H_0} \left(\|\Pi_m y\|_n^2 > q_m^y(\alpha_m) \right) \leq \alpha_m. \quad (1)$$

Ici, nous utilisons la notation $q_m^y(\alpha_m)$ pour insister sur le fait que nous choisirons, au bout du compte, un seuil dépendant de y . Nous rejetons H_0 si l'un au moins de ces tests conduit à rejeter H_0 . Notre région de rejet est donc

$$\sup_{m \in \mathcal{M}} \left\{ \|\Pi_m y\|_n^2 - q_m^y(\alpha_m) \right\} > 0. \quad (2)$$

Il est clair que si les α_m sont choisis tel que $\alpha = \sum_{m \in \mathcal{M}} \alpha_m$, alors notre test est de niveau α . Par exemple, $\alpha_m = \alpha/|\mathcal{M}|$ convient.

1.2 calibration

Il reste à déterminer $q_m^y(\alpha_m)$ vérifiant (1) lorsque m et α_m sont donnés. Si cela était possible, on définirait $q_m^y(\alpha_m)$ comme étant le quantile d'ordre $1 - \alpha_m$ de la loi de $\|\Pi_m y\|_n^2$ sous H_0 , c'est-à-dire la loi de $\|\Pi_m \varepsilon\|_n^2$. Cela est malheureusement impossible sans hypothèses supplémentaires sur la loi de ε . Nous proposons alors de considérer des seuils $q_m^y(\alpha_m)$ aléatoires fondés sur un principe de symétrisation. Précisément, nous définissons

$$q_m^y(\alpha_m) = \inf \left\{ x \in \mathbb{R}, P \left[\|\Pi_m(w \times y)\|_n^2 > x \mid y \right] \leq \alpha_m \right\}, \quad (3)$$

où w est un vecteur aléatoire de \mathbb{R}^n indépendant de y dont les composantes sont i.i.d. et telles que $P(w_i = 1) = P(w_i = -1) = 1/2$, et $w \times y$ désigne le vecteur de \mathbb{R}^n de composantes $w_i y_i$. Ainsi, nous symétrisons y en multipliant chaque composante par un signe aléatoire. Cela n'affecte pas la loi sous H_0 puisque la loi de y est symétrique sous H_0 . Grâce à cela, la contrainte de niveau (1) est bien satisfaite. En revanche, sous H_1 , la symétrisation transforme la loi non symétrique de y en une loi symétrique, imitant le comportement sous l'hypothèse nulle et laissant envisager une bonne puissance. Nous conditionnons par y pour que les seuils $q_m^y(\alpha_m)$ ne dépendent pas de la loi inconnue de y , mais dépendent uniquement de y (qui est bien observé) et de la loi connue de w .

1.3 Implémentation

Plutôt que de faire un calcul exact des seuils $q_m^y(\alpha_m)$ (dont la complexité algorithmique est de l'ordre de 2^n), nous réalisons des simulations de Monte-Carlo de la façon suivante. Nous générons B vecteurs i.i.d. w^1, \dots, w^B de même loi que w , indépendamment de y . Nous considérons alors la région de rejet (2) où $q_m^y(\alpha_m)$ est remplacé, pour tout m , par une version empirique :

$$\widehat{q}_m^B(\alpha_m) = \inf \left\{ x \in \mathbb{R}, \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\|\Pi_m(w^b \times y)\|_n^2 > x} \leq \alpha_m \right\}. \quad (4)$$

2 Performances théoriques

Les Théorèmes 2 et 3 de Durot et Rozenholc (2006) montrent que le niveau et la puissance du test construits avec les seuils définis par (4) sont similaires à ceux du test dont les seuils sont donnés en (3) lorsque $B \rightarrow \infty$. Nous considérons donc seulement le test de région de rejet (2) avec $q_m^y(\alpha_m)$ défini par (3).

2.1 Dans le cadre général

Nous montrons tout d'abord que notre test est de niveau non-asymptotique donné.

Théorème 1. *Nous observons $y = f + \varepsilon$, où $f \in \mathbb{R}^n$ et les ε_i sont des variables aléatoires indépendantes, toutes de loi symétrique. Soit \mathcal{M} une collection de partitions de $\{1, \dots, n\}$. Soient α et $(\alpha_m)_{m \in \mathcal{M}}$ des réels positifs tels que $\alpha = \sum_{m \in \mathcal{M}} \alpha_m$. Alors,*

$$P_{H_0} \left(\sup_{m \in \mathcal{M}} \{ \|\Pi_m y\|_n^2 - q_m^y(\alpha_m) \} > 0 \right) \leq \alpha.$$

Les Théorèmes 4 et 5 de Durot et Rozenholc (2006) établissent des résultats de puissance du test dans le cadre général décrit ci-dessus, sous deux hypothèses de moment différentes. La première hypothèse envisagée est une condition de Bernstein : il existe des réels positifs γ et μ tels que pour tout entier $p \geq 1$,

$$\max_{1 \leq i \leq n} E(\varepsilon_i^{2p}) \leq \gamma p! \mu^{p-2}. \quad (5)$$

La seconde hypothèse envisagée est que les moments sont bornés : il existe $p \geq 2$ et $\mu > 0$ tels que

$$\max_{1 \leq i \leq n} E(\varepsilon_i^{2p}) \leq \mu^p. \quad (6)$$

Baraud *et al.* (2003) sont les rares à avoir considéré un cadre aussi général que le nôtre, mais ils posent l'hypothèse supplémentaire que $\varepsilon_1, \dots, \varepsilon_n$ sont i.i.d. gaussiennes. Ce cas est couvert par nos hypothèses, qui ne supposent pas que $\varepsilon_1, \dots, \varepsilon_n$ sont gaussiennes, ni même que ces variables sont de même loi. La puissance de notre test est comparable à celle du test proposé par Baraud *et al.* (2003) dans le cas Gaussien. Le mérite de notre méthode est d'être performant sous des hypothèses très générales. Considérer des hypothèses générales est rendu possible par le principe de calibration par symétrisation.

2.2 Dans le modèle de régression

Nous nous plaçons dans le cadre du Théorème 1 ci-dessus et supposons de plus que $f_i = F(x_i)$ pour tout $i = 1, \dots, n$, où $x_i \in [0, 1]$ est fixé et $F : [0, 1] \rightarrow \mathbb{R}$ est inconnue. Nous utilisons le test décrit plus haut pour tester

$$H'_0 : F \equiv 0 \quad \text{contre} \quad H'_1 : F \not\equiv 0,$$

en considérant pour \mathcal{M} la collection des partitions dyadiques de $\{1, \dots, n\}$. D'après ce qui précède, ce test est de niveau non-asymptotique α . Pour alléger, nous considérons seulement ici le cas où $x_i = i/n$, $\alpha_m = \alpha/|\mathcal{M}|$, et nous supposons qu'il existe des réels positifs γ et μ tels qu'on ait (5) pour tout entier $p \geq 1$. Pour un réel $\beta \in (0, 1)$ arbitraire, nous donnons dans le corollaire suivant des conditions suffisantes pour que la puissance du test soit au moins $1 - \beta$, c'est-à-dire pour que

$$P_f \left(\sup_{m \in \mathcal{M}} \{ \|\Pi_m y\|_n^2 - q_m^y(\alpha_m) \} > 0 \right) \geq 1 - \beta. \quad (7)$$

Corollaire 1. *Supposons qu'il existe $s \in (1/4, 1]$, $R > 0$ et $L > 0$ tels que*

$$\forall (u, v) \in [0, 1]^2, |F(u) - F(v)| \leq R|u - v|^s \quad \text{et} \quad \sup_{u \in [0, 1]} |F(u)| \leq L. \quad (8)$$

Notons $\delta = \sqrt{\gamma} + \mu + L^2$. Alors, il existe n_0 et C_0 ne dépendant que de α et β tels qu'on ait (7) dès lors que $n \geq n_0$, $C \geq C_0$ et

$$\|F\|_2 \geq CR^{1/(1+4s)} \left(\frac{\delta \sqrt{\log \log n}}{n} \right)^{2s/(1+4s)}.$$

La vitesse de séparation obtenue ici est exactement la vitesse minimax calculée par Spokoiny (1996) dans le modèle de bruit blanc et par Gayraud and Pouet (2005) dans le modèle de régression gaussienne. Ainsi, notre test atteint la vitesse optimale de séparation sous une condition de Bernstein si $s > 1/4$. Il atteint en fait la vitesse optimale de séparation dans des situations plus variées.

2.3 Preuve du Théorème 1.

Cette courte preuve montre comment la symétrisation permet de calibrer un test dans une situation où la loi de la statistique de test n'est pas connue sous H_0 . Sous H_0 , nous avons $y = \varepsilon$ donc nous voulons démontrer que

$$P \left(\sup_{m \in \mathcal{M}} \{ \|\Pi_m \varepsilon\|_n^2 - q_m^\varepsilon(\alpha_m) \} > 0 \right) \leq \alpha. \quad (9)$$

Pour tout $u \in \mathbb{R}^n$, soient $|u|$ et $\text{sgn}(u)$ les vecteurs de \mathbb{R}^n dont les i -ème composantes sont respectivement $|u_i|$ et

$$(\text{sgn}(u))_i = \mathbb{1}_{\{u_i > 0\}} - \mathbb{1}_{\{u_i < 0\}}. \quad (10)$$

La loi de ε est symétrique et

$$w \times \varepsilon = w \times \text{sgn}(\varepsilon) \times |\varepsilon|,$$

donc la loi conditionnelle de $w \times \varepsilon$ sachant ε est identique à la loi conditionnelle de $w \times \varepsilon$ sachant $|\varepsilon|$. En particulier, $q_m^\varepsilon(\alpha_m) = q_m^{|\varepsilon|}(\alpha_m)$, où

$$q_m^{|\varepsilon|}(\alpha_m) = \inf \left\{ x \in \mathbb{R}, P \left[\|\Pi_m(w \times \varepsilon)\|_n^2 > x \mid |\varepsilon| \right] \leq \alpha_m \right\}.$$

Mais conditionnellement à $|\varepsilon|$, $w \times \varepsilon$ a la même loi que $\text{sgn}(\varepsilon) \times |\varepsilon| = \varepsilon$. Donc, pour tout $m \in \mathcal{M}$,

$$P \left(\|\Pi_m \varepsilon\|_n^2 > q_m^\varepsilon(\alpha_m) \mid |\varepsilon| \right) = P \left(\|\Pi_m(w \times \varepsilon)\|_n^2 > q_m^{|\varepsilon|}(\alpha_m) \mid |\varepsilon| \right) \leq \alpha_m.$$

En intégrant l'inégalité précédente, on obtient que

$$P(\|\Pi_m \varepsilon\|_n^2 > q_m^\varepsilon(\alpha_m)) \leq \alpha_m.$$

On en déduit

$$P\left(\sup_{m \in \mathcal{M}} \{\|\Pi_m \varepsilon\|_n^2 - q_m^\varepsilon(\alpha_m)\} > 0\right) \leq \sum_{m \in \mathcal{M}} P(\|\Pi_m \varepsilon\|_n^2 > q_m^\varepsilon(\alpha_m)) \leq \sum_{m \in \mathcal{M}} \alpha_m.$$

Puisque par hypothèse, $\sum_{m \in \mathcal{M}} \alpha_m = \alpha$, on en déduit (9). \square

Bibliographie

- [1] Baraud, Y., Huet, S., & Laurent, B. (2003). Adaptive tests of linear hypotheses by model selection. *Annals of statistics*, 225-251.
- [2] Durot, C., & Rozenholc, Y. (2006). An adaptive test for zero mean. *Mathematical Methods of Statistics*, 15(1), 26-60.
- [3] Gayraud, G., & Pouet, C. (2005). Adaptive minimax testing in the discrete regression scheme. *Probability theory and related fields*, 133(4), 531-558.
- [4] Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6), 2477-2498.