Polynômes locaux et tests d'adéquation conditionnels

Sandie Ferrigno

Laboratoire de Probabilités et Statistique
Département des Sciences Mathématiques
Université Montpellier II
Case Courrier 051-Place Eugène Bataillon
34095 Montpellier Cedex 5
FRANCE
courriel : ferrigno@math.univ-montp2.fr

On considère (X, Y), un couple de variables aléatoires à valeurs dans $\mathbb{R}^2$. Pour étudier la relation liant X à Y, il est courant d'utiliser la "fonction de régression" définie par l'espérance conditionnelle de Y sachant $\{X = x\}$ qui est un résumé pratique de l'effet de X sur le comportement de Y (voir, par exemple, [1] ou [3]).

Dans ce travail, on utilise la fonction de répartition conditionnelle de Y sachant $X = x$, notée

$$F(y|x) = P(Y \leq y|X = x) \tag{1}$$

L'intérêt de cette fonction provient du fait que la plupart des quantités statistiques utilisées en pratique pour comprendre la relation liant X et Y (dont la fonction de régression) en sont des fonctionnelles. Dans de nombreuses procédures statistiques, on émet aussi des hypothèses sur la loi de Y sachant X et il semble donc important de pouvoir les valider à l'aide de tests d'adéquation.

Dans ce travail, on développe une approche globale où toutes les hypothèses faites dans le but d'obtenir un modèle pour $F(y|x)$ sont testées simultanément. Plus précisément, le problème dont nous traitons ici est de construire un test pour valider ou infirmer l'hypothèse nulle :

$$H_0 : F(y|x) = F_0(y|x) \quad \forall(x, y) \tag{2}$$

où $F_0(y|x)$ est la fonction de répartition conditionnelle correspondant au modèle supposé. Si dans le test de (2) tous les paramètres sont connus, $F_0(y|x)$ est dite entièrement spécifiée. Mais généralement ces paramètres sont inconnus. Si on les rassemble dans un vecteur θ appartenant à l'espace paramétrique Θ et si on réécrit la fonction de répartition conditionnelle sous la forme $F(y|x;\theta)$, alors (2) devient l'hypothèse nulle composite

$$H_0 : F(y|x) \in \{F(y|x;\theta), \theta \in \Theta\}. \tag{3}$$

On considère d'abord le problème de tester (2). Pour cela, on propose de comparer $F_0(y|x)$ à un estimateur de $F(y|x)$ et de rejeter H_0 si la "distance" entre $F_0(y|x)$ et cet estimateur est trop grande. Il existe différentes méthodes pour estimer $F(y|x)$ à partir d'un échantillon (X_i, Y_i) , $i = 1 \ldots n$ de n copies indépendantes de (X, Y). On utilise ici une méthode non-paramétrique d'estimation : l'estimation polynômiale locale (voir [2]). Les résultats obtenus dans ce contexte sont ensuite utilisés dans le cas plus pratique où la fonction de répartition conditionnelle contient des paramètres inconnus et pour tester (3).

Enfin, pour permettre la comparaison de nos stratégies de test avec d'autres tests d'adéquation, on propose, dans le cadre d'une analyse de puissance, d'affiner celle-ci en étudiant son comportement asymptotique local.

Bibliographie

[1] Alcalá, J.T., Cristóbal, J.A. & González-Manteiga, W. (1999). Goodness of fit tests for linear models based on local polynomials, *Statistics and Probability letters,* **42**, 39-46.

[2] Fan, J. & Gijbels, I. (1996). Local polynomial modelling and its applications, *Chapman and hall.*

[3] Hàrdle,W. & Mammen, E.(1993). Comparing nonparametric versus parametric regression fits, *The Annals of Statistics,* **21**(4), 39-46.