

ESTIMATION D'UN MODÈLE ADDITIF SUR DONNÉES CENSURÉS PAR LA MÉTHODE DE RÉTRO-AJUSTEMENT DE SPLINES PAR NOYAU OU “SPLINE BACKFITTED KERNEL”.

Samuel Maistre¹, Anouar El Ghouch² & Ingrid Van Keilegom³

¹ *Université Catholique de Louvain, ISBA, Voie du Roman Pays 20 bte L1.04.01
1348 Louvain-la-Neuve, Belgique; samuel.maistre@uclouvain.be*

² *Université Catholique de Louvain; anouar.elghouch@uclouvain.be*

³ *Université Catholique de Louvain; ingrid.vankeilegom@uclouvain.be*

Résumé. Dans le cadre de l'estimation d'un modèle additif non paramétrique, nous présentons l'adaptation de la méthode de “Spline Backfitted Kernel” au cas où la réponse est censurée à droite.

Mots-clés. Modèle additif, statistique non paramétrique, censure.

Abstract. We present the Spline Backfitted Kernel method to estimate a nonparametric additive model in the case of right censoring.

Keywords. Additive model, nonparametric statistics, censoring.

1 La méthode de “Spline Backfitted Kernel”

Les modèles additifs non paramétriques ont été largement étudiés depuis le travail précurseur de Hastie et Tibshirani (1986). Ils permettent un compromis satisfaisant entre une modélisation paramétrique et une modélisation purement non paramétrique. Et puisque chaque fonction d'intérêt est unidimensionnelle, on peut la représenter et donc l'interpréter facilement.

Le modèle additif s'écrit

$$Y = \mu + \sum_{d=1}^D m_d(X_d) + \varepsilon$$

avec $\mathbf{X} = (X_1, \dots, X_D) \in [0, 1]^D$ un vecteur de D covariables et ε une variable centrée conditionnellement à \mathbf{X} . Il s'agit, sur la base d'un échantillon i.i.d. $\{Y_i, \mathbf{X}_i, 1 \leq i \leq n\}$, d'estimer les paramètres μ et $m_d(\cdot)$, $1 \leq d \leq D$. Plusieurs techniques d'estimation ont été proposées, notamment la projection pénalisée sur une base de splines et le rétro-ajustement, notamment par lissage à noyau. Une autre méthode relativement récente (cf. Wang et Yang (2009)) propose de combiner les idées de ces deux méthodes : le rétro-ajustement de splines par noyau ou “Spline Backfitted Kernel”. Nous proposons d'adapter cette méthode au cas où la variable réponse est censurée à droite.

Cette méthode comporte deux étapes :

1. D'abord, on estime le modèle additif en utilisant une projection sur une base de splines dont le nombre de nœuds mène à un sous-lissage. On obtient alors des estimateurs $\hat{\mu}$ et \hat{m}_d , $1 \leq d \leq D$.
2. Ensuite, on utilise les estimateurs de la première étape pour faire un lissage à noyau univarié afin d'estimer chacune des fonctions d'intérêt. Par exemple, pour estimer la première fonction $m_1(\cdot)$, on utilisera l'estimateur final

$$\hat{m}_{1,K}(x_1) = \frac{\sum_{i=1}^n \left(Y_i - \hat{\mu} - \sum_{d=2}^D \hat{m}_d(X_{id}) \right) K_h(x_1 - X_{i1})}{\sum_{i=1}^n K_h(x_1 - X_{i1})}$$

avec $K_h(x) = h^{-1}K(x/h)$, où $K(\cdot)$ est un noyau.

2 Utilisation de l'estimateur SBK lorsque la réponse est censurée à droite

Dans notre contexte, on n'observe pas toujours la réponse mais le couple

$$(T, \delta) = (Y \wedge C, \mathbf{1}\{Y \leq C\})$$

où C est une variable dont on note $G(\cdot)$ la fonction de répartition. On utilise alors l'estimateur de Kaplan-Meier

$$\hat{G}(t) = 1 - \prod_{T_i \leq t} \left(1 - \frac{1}{\sum_{j=1}^n \mathbf{1}_{T_j \geq T_i}} \right)^{1-\delta_i}$$

pour construire des données synthétiques de Koul et al. (1981)

$$\hat{Y}_i = \frac{\delta_i T_i}{1 - \hat{G}(T_i-)}, \quad 1 \leq i \leq n.$$

En effet, \hat{Y}_i est proche de $Y_i^* = \delta_i T_i / (1 - G(T_i-))$ et sous les hypothèses d'indépendance entre Y et C et d'égalité $P(Y \leq C | \mathbf{X}, Y) = P(Y \leq C | Y)$, on a $E[Y^* | \mathbf{X}] = E[Y | \mathbf{X}]$, ce qui permet d'obtenir des résultats similaires au cas non censuré.

En particulier, il est possible de montrer que l'estimation d'une fonction peut être réalisée aussi bien que lorsque l'on connaît tous les autres paramètres inconnus. On parle alors d'efficacité oracle.

Bibliographie

- [1] Hastie, T. and Tibshirani, R. (1986). Generalized additive models *with discussion*. *Statist. Sci.*, **1(3)**, 297–318.

- [2] Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.*, **9(6)**, 1276–1288.
- [3] Wang, J. and Yang, L. (2009). Efficient and fast spline-backfitted kernel smoothing of additive models. *Ann. Inst. Statist. Math.*, **61(3)**, 663–690.