

# Sur la vitesse de convergence de l'estimateur du plus proche voisin baggé

Gérard Biau, Frédéric Cérou, Arnaud Guyader

► **To cite this version:**

Gérard Biau, Frédéric Cérou, Arnaud Guyader. Sur la vitesse de convergence de l'estimateur du plus proche voisin baggé. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. 2009. <inria-00386715>

**HAL Id: inria-00386715**

**<https://hal.inria.fr/inria-00386715>**

Submitted on 22 May 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SUR LA VITESSE DE CONVERGENCE DE L'ESTIMATEUR DU PLUS PROCHE VOISIN BAGGÉ

Gérard BIAU <sup>a</sup>, Frédéric CÉROU <sup>b</sup> et Arnaud GUYADER <sup>b,c</sup>

<sup>a</sup> LSTA & LPMA  
Université Pierre et Marie Curie – Paris VI  
Boîte 158, 175 rue du Chevaleret  
75013 Paris, France  
gerard.biau@upmc.fr

<sup>b</sup> INRIA Rennes Bretagne Atlantique  
Aspi project-team  
Campus de Beaulieu, 35042 Rennes Cedex, France  
Frederic.Cerou@inria.fr

<sup>c</sup> Institut de Recherche Mathématique de Rennes  
UMR CNRS 6625  
Université Rennes II – Haute Bretagne  
Campus Villejean  
Place du Recteur Henri Le Moal, CS 24307  
35043 Rennes Cedex, France  
arnaud.guyader@uhb.fr

**Mots-clés** Apprentissage, Modèles semi et non paramétriques.

**Résumé** On s'intéresse dans cette communication à l'estimation de la fonction de régression  $r(\mathbf{x}) = \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$  associée à un couple aléatoire  $(\mathbf{X}, Y)$  à valeurs dans  $\mathbb{R}^d \times \mathbb{R}$ , à partir d'un échantillon i.i.d.  $\mathcal{D}_n = \{(\mathbf{X}_i, Y_i)_{1 \leq i \leq n}\}$  de même loi que  $(\mathbf{X}, Y)$ . Dans ce contexte, l'estimateur dit du plus proche voisin baggé consiste à tirer dans  $\mathcal{D}_n$  un très grand nombre de sous-échantillons bootstrap indépendants de taille  $k_n < n$ , à considérer pour chacun d'eux l'estimateur du plus proche voisin et enfin à calculer la moyenne de tous ces estimateurs pour prédire. Il a récemment été prouvé par Biau et Devroye (2008) que l'estimateur ainsi obtenu est universellement convergent à condition que  $k_n$  tende vers l'infini et  $k_n/n$  tende vers 0 lorsque  $n \rightarrow \infty$ . Nous montrons dans ce travail que, mieux encore, la vitesse de convergence de cet estimateur baggé est optimale.

**Abstract** Let  $\mathcal{D}_n = \{(\mathbf{X}_i, Y_i)_{1 \leq i \leq n}\}$  be an i.i.d. sample, taking values in  $\mathbb{R}^d \times \mathbb{R}$ , with the same distribution as  $(\mathbf{X}, Y)$ . Our mission is to estimate the regression function  $r(\mathbf{x}) = \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$ . The so-called nearest neighbor bagging consists in the following : draw a huge number of independent subsamples with the same size  $k_n < n$ , calculate, for each subsample, the nearest neighbor estimate and take finally the average of these estimates. It has been recently shown in Biau and Devroye (2008) that this technique yields a universally consistent estimate, provided  $k_n \rightarrow \infty$  and  $k_n/n \rightarrow 0$  as  $n \rightarrow \infty$ . We prove in this communication that the rate of convergence of this estimate is, in fact, optimal.

## 1 Bagging

Les “méthodes d’ensemble” (*ensemble methods* en anglais) sont des algorithmes d’apprentissage qui consistent à entraîner plusieurs prédicteurs que l’on combine pour en obtenir un meilleur. Parmi ces techniques, le bagging (acronyme pour **bootstrap aggregating**) a été proposé par Breiman en 1996. Cette méthode consiste à générer de nombreux sous-échantillons bootstrap des observations initiales, à construire un prédicteur à partir de chacun, et enfin à combiner l’ensemble pour prédire. Il s’agit d’une procédure efficace pour améliorer des estimateurs instables, en particulier pour de grands ensembles de données et en grande dimension. Parmi les contributions théoriques récentes à l’étude du bagging et des méthodes dérivées, on peut citer Buja et Stuetzle (2000a et 2000b), Friedman et Hall (2000), Bühlmann et Yu (2002), Hall et Samworth (2005), ou encore Biau et Devroye (2008).

Dans cette communication, nous examinons les liens entre le principe proposé par Breiman et la méthode de régression dite *du plus proche voisin*. A cette fin, nous supposons disposer d’un échantillon i.i.d.  $\mathcal{D}_n = \{(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)\}$  à valeurs dans  $\mathbb{R}^d \times \mathbb{R}$  de même loi qu’un couple aléatoire générique  $(\mathbf{X}, Y)$  satisfaisant  $\mathbb{E}|Y| < \infty$ . L’espace  $\mathbb{R}^d$  est muni de la métrique euclidienne standard. Pour  $\mathbf{x} \in \mathbb{R}^d$  fixé, notre objectif consiste à estimer la fonction de régression  $r(\mathbf{x}) = \mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$  en utilisant  $\mathcal{D}_n$ . Dans ce contexte, nous dirons qu’un estimateur de la fonction de régression  $r_n(\mathbf{x})$  est convergent lorsque  $\mathbb{E}[r_n(\mathbf{X}) - r(\mathbf{X})]^2 \rightarrow 0$ , lorsque  $n \rightarrow \infty$ . Lorsque cette dernière propriété est vérifiée pour toutes les lois de probabilités de  $(\mathbf{X}, Y)$  telles que  $\mathbb{E}|Y| < \infty$ , on dit en outre que l’estimateur est universellement convergent.

## 2 Bagging et plus proche voisin

Rappelons que l’estimateur du plus proche voisin de la régression s’écrit  $r_n(\mathbf{x}) = Y_{(1)}(\mathbf{x})$  où  $Y_{(1)}(\mathbf{x})$  est l’observation issue du vecteur  $\mathbf{X}_{(1)}(\mathbf{x})$  dont la distance euclidienne à  $\mathbf{x}$  est minimale parmi les observations  $\mathbf{X}_1, \dots, \mathbf{X}_n$ . Il est connu que cette technique ne conduit pas, en général, à un estimateur convergent de la régression (Devroye, Györfi et Lugosi,

1996, Chapitre 5).

Notons  $r_{n,k_n}$  ( $1 \leq k_n \leq n$  est un paramètre) l'estimateur du plus proche voisin pour un sous-échantillon aléatoire de taille  $k_n$  tiré avec (ou sans) remise dans l'échantillon initial  $\mathcal{D}_n$ . La technique bagging consiste alors à répéter ce sous-échantillonnage aléatoire un nombre infini de fois et à effectuer la moyenne des résultats obtenus. Ainsi, d'un point de vue théorique, l'estimateur baggé de la régression a pour expression

$$r_n^*(\mathbf{x}) = \mathbb{E}^* [r_{n,k_n}(\mathbf{x})],$$

où  $\mathbb{E}^*$  désigne l'espérance par rapport au rééchantillonnage, conditionnellement aux données  $\mathcal{D}_n$ . En pratique, le rééchantillonnage est effectué par méthode Monte-Carlo : on répète l'opération  $m$  fois ( $m$  grand) et on fait la moyenne.

Le résultat suivant, prouvé dans Biau et Devroye (2008), montre que pour un choix approprié de  $k_n$ , la version baggée de la régression au plus proche voisin est universellement convergente :

**Theorem 2.1** *Si  $k_n \rightarrow \infty$  et  $k_n/n \rightarrow 0$ , alors l'estimateur  $r_n^*$  est universellement convergent.*

Il est important de mentionner que, dans ce résultat, peu importe que le sous-échantillonnage soit fait avec ou sans remise. On retient donc que la méthode du bagging permet de transformer l'estimateur de plus proche voisin (qui n'est pas convergent en général) en un estimateur convergent, pourvu que la taille des sous-échantillons reste négligeable par rapport à la taille de l'échantillon initial. La preuve du Théorème 2.1 repose sur l'observation suivante : l'estimateur  $r_n^*(\mathbf{x})$  est en fait un estimateur de la forme

$$\sum_{i=1}^n V_i Y_{(i)}(\mathbf{x}),$$

avec

$$V_i = \mathbb{P}(\text{le } i\text{-ème plus proche voisin de } \mathbf{x} \text{ est le plus proche voisin du sous-échantillon}).$$

Un tel estimateur porte le nom d'estimateur des plus proches voisins pondérés (Stone, 1977, et Problèmes 11.7, 11.8 de Devroye et al., 1996).

Dans le cas "sans remise", on voit facilement que les  $V_i$  ont pour expression

$$V_i = \begin{cases} \frac{\binom{n-i}{k_n-1}}{\binom{n}{k_n}}, & i \leq n - k_n + 1 \\ 0, & i > n - k_n + 1. \end{cases} \quad (2.1)$$

tandis que dans le cas “avec remise” on a

$$V_i = \left(1 - \frac{i-1}{n}\right)^{k_n} - \left(1 - \frac{i}{n}\right)^{k_n}. \quad (2.2)$$

Forts de cette constatation, nous étudions dans cette communication les vitesses de convergence de  $\mathbb{E} [r_n^*(\mathbf{x}) - r(\mathbf{x})]^2$  lorsque  $n \rightarrow \infty$ . Notre principal résultat est le suivant :

**Theorem 2.2** *Supposons que le support de  $\mathbf{X}$  est borné et que la fonction de régression  $r$  est lipschitzienne. Alors, si  $d \geq 3$ ,*

$$\mathbb{E} [r_n^*(\mathbf{x}) - r(\mathbf{x})]^2 = O\left(n^{-\frac{2}{d+2}}\right)$$

avec le choix  $k_n = n^{\frac{d}{d+2}}$ .

Nous présenterons également des résultats pour les cas particuliers  $d = 1$  et  $d = 2$ .

## Bibliographie

- [1] Biau, G. and Devroye, L. (2008). On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification, *Technical Report, Université Pierre et Marie Curie, Paris*.
- [2] Breiman, L. (1996). Bagging predictors, *Machine Learning*, **24**, 123-140.
- [3] Bühlmann, P. and Yu, B. (2002). Analyzing bagging, *The Annals of Statistics*, **30**, 927-961.
- [4] Buja, A. and Stuetzle, W. (2000a). The effect of bagging on variance, bias, and mean squared error, *Technical Report, University of Pennsylvania, Philadelphia*.
- [5] Buja, A. and Stuetzle, W. (2000b). Smoothing effects of bagging, *Technical Report, University of Pennsylvania, Philadelphia*.
- [6] Devroye, L., Györfi, L. and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*, Springer-Verlag, New York.
- [7] Friedman, J.H. and Hall, P. (2000). On bagging and nonlinear estimation, *Technical Report, Stanford University, Stanford*.
- [8] Hall, P. and Samworth, R.J. (2005). Properties of bagged nearest neighbour classifiers, *Journal of the Royal Statistical Society B*, **67**, 363-379.
- [9] Stone, C.J. (1977). Consistent nonparametric regression, *The Annals of Statistics*, **5**, 595-645.