Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 343 (2006) 665–669
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics/Probability Theory

Almost sure convergence of the k_T-occupation time density estimator

Boris Labrador

L.S.T.A., université Paris 6, 175, rue du Chevaleret, 75013 Paris, France

Received 10 February 2006; accepted after revision 10 October 2006

Presented by Paul Deheuvels

Abstract

In this Note, we introduce an extension of the k-nearest neighbor estimator in continuous time, the k_T-occupation time estimator, and we give sufficient conditions for its existence. Then, we show the almost sure convergence for α-mixing and bounded processes in two cases, the superoptimal case (when parametric rates are reached) and the optimal case (when i.i.d. rates of density estimation are reached). ***To cite this article: B. Labrador, C. R. Acad. Sci. Paris, Ser. I 343 (2006).***

Résumé

Convergence presque sûre de l'estimateur de la densité du k_T-temps d'occupation. Dans cette Note, nous introduisons une extension de l'estimateur des k-plus proches voisins en temps continu, l'estimateur du k_T-temps d'occupation, puis nous donnons des conditions d'existence de cet estimateur. Nous établissons également la convergence presque sûre pour des processus bornés α-mélangeants dans deux cas, le cas suroptimal (où la vitesse paramétrique est atteinte) et le cas optimal (où la vitesse i.i.d. de l'estimation de la densité est atteinte). ***Pour citer cet article : B. Labrador, C. R. Acad. Sci. Paris, Ser. I 343 (2006).***

Version française abrégée

L'estimation non-paramétrique de la densité lorsque la trajectoire est observée sur [0,T] avec $T \to \infty$ a fait l'objet de nombreuses études (voir par exemple Bosq [2], Kutoyants [11] et les références incluses). Dans cette Note, nous allons introduire une extension de l'estimateur des k-plus proches voisins en temps continu, que nous appellerons estimateur du k_T-temps d'occupation, puis nous nous intéresserons à la convergence presque sûre de cet estimateur dans le cas optimal et dans le cas suroptimal (voir Bosq [2]).

La méthode des k-plus proches voisins a d'abord été introduite par Fix et Hodges [9] en 1951 puis, en 1965, Loftsgaarden et Quesenberry [12] ont étudié la convergence simple en probabilité de l'estimateur des k plus proches voisins basé sur l'échantillon $(X_1, \ldots, X_n)$. Pour constuire cet estimateur, nous choisissons un entier $k(n)$ que l'on divise par n fois le volume de l'hypersphère de centre x contenant $k(n)$ observations dans $\mathbb{R}^p$. Dans le cas i.i.d., les proprié-

E-mail address: labrador@ccr.jussieu.fr (B. Labrador).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2006.10.015

tés asymptotiques de cet estimateur ont été abondamment étudiées. En particulier, pour la convergence presque sûre ponctuelle ou uniforme, nous pouvons citer les travaux de Devroye et Penrod [6], Devroye et Wagner [7], Moore et Henrichon [14], Moore et Yackel [15], Wagner [19]. Des vitesses précises sont également obtenues par Chen [4], Deheuvels et Mason [5], Mack [13], Ralescu [16].

Dans le cadre dépendant, nous pouvons citer Boente et Fraiman [1] qui ont obtenu des vitesses de convergence uniforme en supposant les processus α-mélangeants ou φ-mélangeants. Sous une condition de mélangeance plus générale que la forte mélangeance (voir par exemple [8]), Tran [18] établit la convergence presque complète de l'estimateur. Nous nous proposons d'étendre les résultats de Tran en temps continu en les complétant avec des vitesses de convergence. Pour cela, nous définissons une extension de l'estimateur des plus proches voisins en temps continu, l'estimateur du k_T-temps d'occupation. Soit k_T donné ($0 < k_T < T$), nous introduisons $\delta_T := \delta_{T,x}$ un rayon aléatoire tel que le processus passe un temps k_T dans l'intervalle fermé I_{x,δ_T} de centre x et de rayon δ_T. Nous pouvons écrire $k_T = \int_{t=0}^{T} \mathbf{1}_{I_{x,\delta_T}}(X_t)\,\mathrm{d}t$ où X_t est un processus à valeurs réelles. L'estimateur du k_T-temps d'occupation est donné par

$$f_T(x) = \frac{k_T}{2T\delta_T}.$$

L'existence est assurée par l'hypothèse du temps local (cf. Geman et Horowitz [10] pour les conditions d'existence et les propriétés du temps local).

Pour obtenir les résultats de convergence, nous utilisons les hypothèses classiques de l'estimateur des k-plus proches voisins $k_T \to \infty$, $\frac{k_T}{T} \to 0$ avec une hypothèse de régularité sur la dérivée première de f. Le processus est supposé mélangeant borné avec une hypothèse faible de stationnarité. Afin de spécifier les vitesses de convergence, nous reprenons des hypothèses usuelles du temps continu. Pour des processus à trajectoires irrégulières (vérifiant la condition de Castellana–Leadbetter [3] satisfaite en particulier par des processus de diffusions ergodiques, voir Kutoyants [11]), une vitesse suroptimale de convergence (de l'ordre de $\frac{\ln T}{\sqrt{T}}$) est obtenue. Pour étudier une plus grande classe de processus, nous reprenons les hypothèses introduites par Bosq sous lesquelles nous retrouvons, à un terme logarithmique près, la vitesse obtenue dans le cas i.i.d. (voir Ralescu [16], Zhang [20]). Plus précisément, pour tout $x \in \mathbb{R}$ et pour toute suite $(T_n)_{n\in\mathbb{N}}$, $T_n \uparrow \infty$, $T_{n+1} - T_n \geqslant d > 0$, nous avons le résultat suivant

$$\limsup_n \varepsilon_{T_n}^{-1} \left| f_{T_n}(x) - f(x) \right| < \infty \quad \text{p.s.},$$

où la vitesse ε_{T_n} dépend du choix de k_{T_n} selon le cas considéré : $\varepsilon_{T_n} = \frac{\ln T_n}{\sqrt{T_n}}$ (dans le cas suroptimal) et $\varepsilon_{T_n} = \left(\frac{\ln T_n}{T_n}\right)^{1/3}$ (dans le cas optimal).

1. Introduction

The method of nearest neighbor was initially proposed by Fix and Hodges [9] in 1951. And, in the setting of nonparametric estimation of the density, the k-nearest neighbor estimator (the k-NN estimator) was introduced and studied in the i.i.d. case by Loftsgaarden and Quesenberry [12] in 1965. The k-NN estimator has random bandwidth, that is to say, to estimate the pointwise density f, we fix the number of observations in a sample $(X_1, \ldots, X_n)$ and the measure of the neighborhood of x is then randomized. In literature, we can find numerous results of convergence (essentially pointwise and uniform convergence), for example, Devroye and Penrod [6], Devroye and Wagner [7], Moore and Henrichon [14], Moore and Yackel [15], Wagner [19] among others. Moreover, accurate rates are obtained by Chen [4], Deheuvels and Mason [5], Mack [13], Ralescu [16]. In the dependent case, rates have been exhibited also, we may refer to Boente and Fraiman [1] among others.

In continuous time, nonparametric density estimation, when a sample path is observed over $[0, T]$, has received special attention (see e.g. Bosq [2], Kutoyants [11] and references therein). In this Note, the aim is to obtain pointwise convergence of the density by an extension of the k-NN estimator in continuous time with strong mixing conditions for bounded processes. We call this new estimator the k_T-Occupation Time estimator (the k_T-OT estimator). As in the discrete case, the bandwidth is randomized but here k_T represents the time spent in the interval $[0, T]$. In order to obtain good properties of this estimator, we need two antagonist relations, k_T should be large, on the one hand, and, on the other hand, $\frac{k_T}{T}$ should be small. In this Note, we extend a result from Tran [18] in continuous time and add rates of convergence. Actually, we can show, in the superoptimal case and in the optimal case, the following consistency result: for all $x \in \mathbb{R}$, $\limsup_n \varepsilon_{T_n}^{-1} |f_{T_n}(x) - f(x)| < \infty$ *a.s.* for any sequence $(T_n)_{n\in\mathbb{N}}$, $T_n \uparrow \infty$, $T_{n+1} - T_n \geqslant d > 0$ where f is strictly positive with continuous derivative in a neighborhood of the point x on $\mathbb{R}$ and ε_{T_n} is the rate depending on the choice of k_{T_n}.

1.1. Notations and assumptions

Let $\mathcal{X} = \{X_t, t \in \mathbb{R}\}$ be a real-valued continuous time process defined and measurable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Moreover, we may suppose $\mathcal{X}$ with continuous sample path, observed over $[0, T]$ with local time. For the existence and properties of local time, we refer to Geman and Horowitz [10]. We may also suppose that the X_t's have a common distribution μ with an unknown density f and that the random vector (X_s, X_t) has a density for all (s, t), $s \neq t$. We set $g_{s,t} := f_{(X_s,X_t)} - f \otimes f$, for all $s \neq t$ and note $I_{x,r}$ the closed interval centered at x and of radius r.

To measure the dependence between the observed variables, we use the strong mixing coefficient (see Doukhan [8]). In order to obtain the convergence of the k_T-OT estimator, we will have the following hypotheses H_1, $\mathrm{H}_{2.j}$, $\mathrm{H}_{3.j}$ with $j = 1$ in the superoptimal case and $j = 2$ in the optimal one respectively given by

H_1:
- (i) $k_T \to \infty$, $\frac{k_T}{T} \to 0$,
- (ii) for all fixed x in $\mathbb{R}$, the density f is strictly positive and has continuous derivative in a neighborhood of the point x,
- (iii) $\{X_t, t \in \mathbb{R}\}$ is a bounded continuous time process: $|X_t| \leqslant M$ where $M < \infty$. Moreover, the stationary condition $g_{|t-s|} := g_{s,t}$ for all $s, t \in \mathbb{R}$ is fulfilled.

$\mathrm{H}_{2.1}$: $\int_0^\infty \|g_u\|_\infty \, \mathrm{d}u < \infty$ (Castellana–Leadbetter's Condition [3])

or

$\mathrm{H}_{2.2}$: There exists $\Gamma \in \mathcal{B}_{\mathbb{R}^2}$ containing the diagonal $D = \{(s,t) \in \mathbb{R}^2, s = t\}$ such as
- (i) $\sup_{s,t \notin \Gamma} \|g_{s,t}\|_\infty < \infty$,
- (ii) $l_{\Gamma,T} = \frac{1}{T} \int_{[0,T]^2 \cap \Gamma} \mathrm{d}s \, \mathrm{d}t$ with $\limsup_{T\to\infty} l_{\Gamma,T} = l_\Gamma < \infty$.

$\mathrm{H}_{3.1}$: $\{X_t,\ t \in \mathbb{R}\}$ is a geometrically strong mixing process: $\alpha(u) \leqslant a\rho^u$ where $a > 0$, $0 < \rho < 1$

or

$\mathrm{H}_{3.2}$: $\{X_t, t \in \mathbb{R}\}$ is an arithmetically strong mixing process: $\alpha(u) \leqslant bu^{-\beta}$ where $\beta > 5$ and $b > 0$.

Remark 1. In the i.i.d. case, H_1(i) is a necessary and sufficient condition for the almost sure convergence of the estimator (see S. Levallois [17], Th. 1.4.1 p. 11). In continuous time, under α-mixing hypothesis, H_1(i)–(ii) with the condition $\alpha(u) \leqslant bu^{-\beta}$ where $\beta > 0$ and $b > 0$ are sufficient conditions for the almost sure convergence of the estimator.

$\mathrm{H}_{2.1}$ may be related to irregular sample path properties of the process, it is more restrictive than $\mathrm{H}_{2.2}$. Such condition is particularly fulfilled by ergodic processes (see Kutoyants [11]). For a large class of processes, we use, the hypothesis $\mathrm{H}_{2.2}$ introduced by Bosq [2] to obtain similar rates as in the i.i.d. case (see Ralescu [16] and Zhang [20]).

1.2. Definition and existence of the k_T-occupation time estimator

1.2.1. Definition of the k_T-OT estimator

Let a given k_T $(0 < k_T < T)$, we define a radius $\delta_T := \delta_{T,x}$ such as k_T represents the time spent in the closed interval $[0, T]$. The radius δ_T is a random variable dependent on x and on the sample path. We note that $k_T = \int_{t=0}^T \mathbf{1}_{I_{x,\delta_T}}(X_t)\,\mathrm{d}t$. Now, we define the k_T-OT estimator in continuous time by

$$f_T(x) = \frac{k_T}{2T\delta_T}. \tag{1}$$

1.2.2. Existence of the k_T-OT estimator

To begin with, let us consider the occupation measure ν_T of the process $\mathcal{X}$ defined by $\nu_T(I) = \int_{t=0}^{T} \mathbf{1}_I(X_t)\,\mathrm{d}t$, $I \in \mathcal{B}_{\mathbb{R}}$. If ν_T is a.s. absolutely continuous with respect to Lebesgue measure, then local time for $\mathcal{X}$ is defined as a measurable random function $l_T(x, w)$ such as $l_T(\cdot, w)$ is a version of $\frac{\mathrm{d}\nu_T}{\mathrm{d}\lambda}$ for almost all ω in Ω.

By the definition of local time, we have $\int_{t=0}^{T} \mathbf{1}_I(X_t)\,\mathrm{d}t = \int_I l_T(u)\,\mathrm{d}u$, $I \in \mathcal{B}_{\mathbb{R}}$. Then, one may easily show that $J(\delta) = \int_{t=0}^{T} \mathbf{1}_{I_{x,\delta}}(X_t)\,\mathrm{d}t$ is a continuous and increasing function over $[0, T]$ so that $J(0) = 0$ and $J(\delta) \to T$ as $\delta \to \infty$. Now, if $k_T \in]0, T[$, there exists at least one δ_T so that $k_T = \int_{t=0}^{T} \mathbf{1}_{I_{x,\delta_T}}(X_t)\,\mathrm{d}t$ is well defined. And the uniqueness (in the case of absolutely continuous sample paths) is ensured since the existence of local time implies in particular that $P(X'_t = 0) = 0$ for almost all $t \in [0, T]$ (cf. Geman and Horowitz [10]).

2. Results

We obtain the following theorems:

Theorem 2.1. *Let the assumptions* H_1, $\mathrm{H}_{2.1}$ *and* $\mathrm{H}_{3.1}$ *be fulfilled.*
Then, for any sequence $(T_n)_{n\in\mathbb{N}}$, $T_n \nearrow^{\infty}$, $T_{n+1} - T_n \geqslant d > 0$, *we set*

$$k_{T_n} = (\ln T_n)\sqrt{T_n} \tag{2}$$

and therefore, we have for all $x \in \mathbb{R}$:

$$\limsup_n \frac{\sqrt{T_n}}{\ln T_n}\left|f_{T_n}(x) - f(x)\right| < \infty \quad a.s. \tag{3}$$

Remark 2. The proof gives an explicit upper bound $C(x)$ depending on ρ, $f(x)$, $f'(x)$.

Theorem 2.2. *Let the assumptions* H_1, $\mathrm{H}_{2.2}$ *and* $\mathrm{H}_{3.2}$ *be fulfilled.*
Then, for any sequence $(T_n)_{n\in\mathbb{N}}$, $T_n \nearrow^{\infty}$, $T_{n+1} - T_n \geqslant d > 0$, *we set*

$$k_{T_n} = (\ln T_n)^{1/3} T_n^{2/3} \tag{4}$$

and therefore, we have for all $x \in \mathbb{R}$:

$$\limsup_n \left(\frac{T_n}{\ln T_n}\right)^{1/3} \left|f_{T_n}(x) - f(x)\right| < \infty \quad a.s. \tag{5}$$

Remark 3. The proof gives an explicit upper bound $D(x)$ depending on l_Γ, $f(x)$, $f'(x)$.

Sketch of the proofs. First, one has the decomposition

$$\left\{\left|f_T(x) - f(x)\right| \geqslant \varepsilon_T\right\} = A_T \cup B_T$$

where

$$A_T = \left\{\delta_T \leqslant \frac{k_T}{2T(f(x) + \varepsilon_T)}\right\}$$

and

$$\begin{cases} B_T = \left\{\delta_T \geqslant \dfrac{k_T}{2T(f(x) - \varepsilon_T)}\right\} & \text{if } f(x) > \varepsilon_T, \\ B_T = \emptyset & \text{if } f(x) \leqslant \varepsilon_T. \end{cases}$$

Then, we set

$$\varepsilon_T = M_1 \frac{\ln T}{\sqrt{T}}, \quad M_1 < \infty$$

in the superoptimal case and

$$\varepsilon_T = M_2\left(\frac{\ln T}{T}\right)^{1/3}, \quad M_2 < \infty$$

in the optimal case.

In order to treat $P(A_T)$ and $P(B_T)$ respectively, we use the blocking argument to create blocks from the events A_T and B_T respectively. Then, we show that the two events A_T and B_T can be handled similarly. We use a coupling result to approximate the blocks by independent ones. This way, we may apply an exponential inequality, on the one hand, and, on the other hand, a coupling result to upper bound these two probabilities. Finally, suitable choices of k_T allow us to conclude with the Borel–Cantelli lemma. □

Acknowledgements

I am entirely grateful to D. Blanke for carefully reading this Note. I would like to thank Professor D. Bosq for his advice and the two referees for helpful suggestions.

References

[1] G. Boente, R. Fraiman, Strong order of convergence and asymptotic distribution of nearest neighbor density estimates from dependent observations, Sankhyā Ser. A 53 (2) (1991) 194–205.

[2] D. Bosq, Nonparametric Statistics for Stochastic Processes. Estimation and Prediction, second ed., Lecture Notes in Statist., vol. 110, Springer-Verlag, New York, 1998.

[3] J.V. Castellana, M.R. Leadbetter, On smoothed probability density estimation for stationary processes, Stochastic Process. Appl. 21 (1986) 179–193.

[4] X.-R. Chen, Convergence rates of nearest neighbor density estimates, Sci. Sinica Ser. A 25 (5) (1982) 455–467.

[5] P. Deheuvels, D.M. Mason, Functional laws of the iterated logarithm for the increments of empirical and quantile processes, Ann. Probab. 20 (3) (1992) 1248–1287.

[6] L. Devroye, C.S. Penrod, The strong uniform convergence of multivariate variable kernel estimates, Canad. J. Statist. 14 (3) (1986) 211–219.

[7] L.P. Devroye, T.J. Wagner, The strong uniform consistency of nearest neighbor density estimates, Ann. Statist. 5 (3) (1977) 536–540.

[8] P. Doukhan, Mixing. Properties and Examples, Lecture Notes in Statist., vol. 85, Springer-Verlag, New York, 1994.

[9] E. Fix, J.L. Hodges Jr., Discriminatory analysis, nonparametric discrimination: consistency properties, Report N$^{\text{O}}$ 4, project No 21-49-004, USAF School of Aviation Medicine, Randolph Field, TX.

[10] D. Geman, J. Horowitz, Occupation densities, Ann. Probab. 8 (1) (1980) 1–67.

[11] Y.A. Kutoyants, Statistical Inference for Ergodic Diffusion Processes, Springer Ser. Statist., Springer, 2003.

[12] D.O. Loftsgaarden, C.P. Quesenberry, A nonparametric estimate of a multivariate density function, Ann. Math. Statist. 36 (1965) 1049–1051.

[13] Y.P. Mack, Rate of strong uniform convergence of k-NN density estimates, J. Statist. Plann. Inference 8 (2) (1983) 185–192.

[14] D.S. Moore, E.G. Henrichon, Uniform consistency of some estimates of a density function, Ann. Math. Statist. 40 (1969) 1499–1502.

[15] D.S. Moore, J.W. Yackel, Consistency properties of nearest neighbor density function estimators, Ann. Statist. 5 (1) (1977) 143–154.

[16] S.S. Ralescu, The law of the iterated logarithm for the multivariate nearest neighbor density estimators, J. Multivariate Anal. 53 (1) (1995) 159–179.

[17] S. Levallois, PhD Thesis, University of Montpellier II, 1998.

[18] L.T. Tran, On multivariate variable-kernel density estimates for time series, Canad. J. Statist. 19 (4) (1991) 371–387.

[19] T.J. Wagner, Strong consistency of nonparametric estimate of a density function, IEEE Trans. Systems Man Cybernet. 3 (1973) 289–290.

[20] D.-X. Zhang, Rates of strong uniform convergence of nearest neighbor density estimates on any compact set, Appl. Math. Mech. 11 (4) (1990) 385–393 (English ed.).