C. R. Acad. Sci. Paris, Ser. I 348 (2010) 317–321

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

Uniform in bandwidth consistency of the kernel-type estimator of the Shannon's entropy

Loi du logarithme uniforme pour un estimateur non paramétrique de l'entropie de Shannon

Salim Bouzebda, Issam Elhattab

L.S.T.A., Université Pierre et Marie Curie-Paris 6, 175, rue du Chevaleret, 8ème étage, bâtiment A, 75013 Paris, France

ARTICLE INFO

Article history:
Received 22 July 2008
Accepted after revision 2 December 2009
Available online 25 February 2010

Presented by Paul Deheuvels

ABSTRACT

We establish uniform-in-bandwidth consistency for kernel-type estimators of the differential entropy. Our proofs rely on the methods of Einmahl and Mason (2005) [10].

RÉSUMÉ

Dans cette Note, nous obtenons la consistance uniforme en terme de la fenêtre pour l'estimateur non paramétrique de l'entropie. Nos arguments de démonstration sont basés sur les résultats obtenus par Einmahl et Mason (2005) [10].

Version française abrégée

Soit $(\mathbf{X}_n)_{n\geqslant 1}$ une suite de vecteurs aléatoires indépendants et identiquement distribués (i.i.d.), à valeurs dans $\mathbb{R}^d$, $d\geqslant 1$, ayant $\mathbb{F}$ comme fonction de répartition et f comme densité de probabilité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. L'entropie associée à la fonction f est définie par $H(f)=-\mathbb{E}\{\log(f(\mathbf{X}))\}$, lorsque cette espérance est finie. L'estimateur à noyau de la densité $f(\mathbf{x})$ est défini, pour tout $\mathbf{x}\in\mathbb{R}^d$, par

$$f_{n,h_n}(\mathbf{x})=(nh_n)^{-1}\sum_{i=1}^{n}K\big((\mathbf{x}-\mathbf{X}_i)/h_n^{1/d}\big),$$

où K est une fonction mesurable vérifiant les conditions (K.1-2-3-4) données ci-dessous et $(h_n)_{n\geqslant 1}$ est une suite de réels positifs tels que $0<h_n\leqslant 1$.

Étant donné l'estimateur f_{n,h_n} de f, la méthode *plug-in* nous permet d'estimer la quantité $H(f)$, pour tout $n\geqslant 2$, par

$$H_{n,h_n,\beta}(f)=-\int\limits_{A_{n,\beta}} f_{n,h_n}(\mathbf{x})\log\big(f_{n,h_n}(\mathbf{x})\big)\,\mathrm{d}\mathbf{x},$$

où $A_{n,\beta}:=\{\mathbf{x}\colon\ f_{n,h_n}(\mathbf{x})\geqslant(\log n)^{-\beta}\}$, avec une constante $\beta>0$. Pour des considérations de facilités techniques, le paramètre de centrage utilisé dans cette Note est défini, pour tout $n\geqslant 2$, par

E-mail addresses: salim.bouzebda@upmc.fr (S. Bouzebda), issam.elhattab@upmc.fr (I. Elhattab).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2009.12.007

$$\widehat{\mathbb{E}}H_{n,h_n\beta} = -\int_{A_{n,\beta}} \mathbb{E}f_{n,h_n}(\mathbf{x})\log\big(\mathbb{E}f_{n,h_n}(\mathbf{x})\big)\,\mathrm{d}\mathbf{x}.$$

Les principaux résultats obtenus dans ce travail sont les suivants :

(i) Sous les conditions du Théorème 2.1, pour tout $\beta > 0$, on a, presque sûrement,

$$\limsup_{n\to\infty} \sup_{cn^{-1}(\log n)^{1+4\beta}\leqslant h\leqslant 1} \frac{\sqrt{nh}|H_{n,h,\beta}(f) - \widehat{\mathbb{E}}H_{n,h,\beta}(f)|}{\sqrt{\{\log n\}^{4\beta}(\log(1/h)\vee\log\log n)}} < \infty.$$

(ii) Sous les conditions du Corollaire 2.2, pour tout $\beta > 0$, on a, presque sûrement,

$$\lim_{n\to\infty} \sup_{h_n'\leqslant h\leqslant h_n''} \big|H_{n,h,\beta}(f) - H(f)\big| = 0.$$

(iii) Sous les conditions du Théorème 2.3, pour tout $\beta > 0$, on a, presque sûrement,

$$\limsup_{n\to\infty} \sup_{h_n'\leqslant h\leqslant h_n''} \frac{\sqrt{nh}|H_{n,h,\beta}(f) - H(f)|}{\sqrt{\{\log n\}^{4\beta}(\log(1/h)\vee\log\log n)}} < \infty.$$

1. Introduction and estimation

Let $(\mathbf{X}_n)_{n\geqslant 1}$ be a sequence of i.i.d. $\mathbb{R}^d$-valued random vectors, $d \geqslant 1$, with distribution function $\mathbb{F}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leqslant \mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^d$. We set here $\mathbf{X} = (X_1, \ldots, X_d) \leqslant \mathbf{x} = (x_1, \ldots, x_d)$ whenever $X_i \leqslant x_i$, for all $i = 1, \ldots, d$. Assume that $\mathbb{F}$ has a density function f with respect to Lebesgue measure on $\mathbb{R}^d$. The differential entropy of f is defined by the quantity

$$H(f) = -\int_{\mathbb{R}^d} f(\mathbf{x})\log\big(f(\mathbf{x})\big)\,\mathrm{d}\mathbf{x}, \tag{1}$$

whenever this integral is meaningful, and where d$\mathbf{x}$ denotes Lebesgue measure in $\mathbb{R}^d$. We will use the convention that $0\log(0) = 0$ since $u\log(u) \to 0$ as $u \to 0$. The differential entropy concept was introduced by [16]. Since then and because of numerous potential applications, the subject has received a considerable interest handling various problems as estimating the quantity $H(f)$. We refer to [2], and the references therein, for details. The main purpose of the present Note is to establish uniform in bandwidth consistency of the so-called kernel estimator of the entropy functional $H(f)$.

As a first step of our study, we gather hereafter hypotheses needed to establish our results.

(F.1) The functional $H(f)$ is well-defined by (1), in the sense that

$$\big|H(f)\big| < \infty. \tag{2}$$

We refer to [12] for conditions characterizing (2) in terms of f. A kernel K will be any measurable function fulfilling the following conditions:

(K.1) K is of bounded variation on $\mathbb{R}^d$;
(K.2) K is right continuous on $\mathbb{R}^d$, i.e., for any $\mathbf{t} = (t_1, \ldots, t_d)$, we have

$$K(t_1, \ldots, t_d) = \lim_{\varepsilon_1\downarrow 0,\ldots,\varepsilon_d\downarrow 0} K(t_1+\varepsilon_1, \ldots, t_d+\varepsilon_d);$$

(K.3) $\|K\|_\infty := \sup_{\mathbf{t}\in\mathbb{R}^d} |K(\mathbf{t})| =: \kappa < \infty$;
(K.4) $\int_{\mathbb{R}^d} K(\mathbf{t})\,\mathrm{d}\mathbf{t} = 1$.

The well-known Akaike–Parzen–Rosenblatt (refer to [1,13] and [15]) kernel estimator of f is defined, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$f_{n,h_n}(\mathbf{x}) = (nh_n)^{-1}\sum_{i=1}^{n} K\big((\mathbf{x}-\mathbf{X}_i)/h_n^{1/d}\big), \tag{3}$$

where $0 < h_n \leqslant 1$ is the smoothing parameter.

In a second step, given f_{n,h_n} and recall (1), we estimate $H(f)$, for any $n \geqslant 2$, by setting

$$H_{n,h_n,\beta}(f) = -\int_{A_{n,\beta}} f_{n,h_n}(\mathbf{x})\log\big(f_{n,h_n}(\mathbf{x})\big)\,\mathrm{d}\mathbf{x}, \tag{4}$$

where $A_{n,\beta} := \{\mathbf{x}\colon\ f_{n,h_n}(\mathbf{x}) \geqslant (\log n)^{-\beta}\}$. Here, $\beta > 0$ is a specified constant.

The limiting behavior of f_{n,h_n}, for appropriate choices of the bandwidth h_n, has been extensively investigated in the literature (refer to [3,6], [7] and [14]). In particular, under our assumptions, the condition that $h_n \to 0$ together with $nh_n \to \infty$ is necessary and sufficient for the convergence in probability of $f_{n,h_n}(\mathbf{x})$ towards the limit $f(\mathbf{x})$, independently of $\mathbf{x} \in \mathbb{R}^d$ and the density f. Various uniform consistency results involving the estimator f_{n,h_n} have been recently established. We refer to [5,10], [4] and [9]. In this Note we will use their methods to establish convergence results for the estimate $H_{n,h_n,\beta}(f)$ of $H(f)$. As the central parameter, we consider, for any $n \geqslant 2$, the quantity

$$\widehat{\mathbb{E}}H_{n,h_n,\beta}(f) = -\int_{A_{n,\beta}} \mathbb{E}f_{n,h_n}(\mathbf{x}) \log\big(\mathbb{E}f_{n,h_n}(\mathbf{x})\big)\,\mathrm{d}\mathbf{x},$$

which is more computationally convenient that the usual mathematical expectation $\mathbb{E}H_{n,h_n,\beta}(f)$.

The remainder of this Note is organized as follows. In Section 2, we state our main results concerning the limiting behavior of $H_{n,h_n,\beta}(f)$. The sketch of the proofs of our results is postponed until Section 3.

2. Main results

The main result is given in the following theorem:

Theorem 2.1. *Let K satisfy* (K.1-2-3-4)*, and let f be a bounded density fulfill* (F.1)*. Then for any $\beta > 0$, there exists a function $\Upsilon(c)$ of $c > 0$, such that, for any $c > 0$, we have*

$$\limsup_{n\to\infty} \sup_{cn^{-1}(\log n)^{1+4\beta} \leqslant h \leqslant 1} \frac{\sqrt{nh}|H_{n,h,\beta}(f) - \widehat{\mathbb{E}}H_{n,h,\beta}(f)|}{\sqrt{\{\log n\}^{4\beta}(\log(1/h) \vee \log\log n)}} \leqslant \Upsilon(c) \quad \textit{a.s.}$$

Let $(h'_n)_{n\geqslant 1}$ and $(h''_n)_{n\geqslant 1}$ be two sequences of constants such that $0 < h'_n < h''_n < 1$, together with $h''_n \to 0$ and for any $\beta > 0$, $nh'_n/(\log n)^{1+4\beta} \to \infty$, as $n \to \infty$. A direct application of Theorem 2.1 shows that, with probability 1,

$$\sup_{h'_n \leqslant h \leqslant h''_n} \big|H_{n,h,\beta}(f) - \widehat{\mathbb{E}}H_{n,h,\beta}(f)\big| = O\left(\sqrt{\frac{\{\log n\}^{4\beta}(\log(1/h'_n) \vee \log\log n)}{nh'_n}}\right).$$

This, in turn, implies that

$$\lim_{n\to\infty} \sup_{h'_n \leqslant h \leqslant h''_n} \big|H_{n,h,\beta}(f) - \widehat{\mathbb{E}}H_{n,h,\beta}(f)\big| = 0 \quad \text{a.s.} \tag{5}$$

The following result handles the uniform deviation of the estimate: $H_{n,h_n,\beta}(f)$ with respect to $H(f)$:

Corollary 2.2. *Let K satisfy* (K.1-2-3-4)*, and let f be a uniformly Lipschitz continuous and bounded density on $\mathbb{R}^d$, fulfilling* (F.1)*. Then for any $\beta > 0$, and for each pair of sequences $0 < h'_n < h''_n \leqslant 1$ with $h''_n \to 0$, $nh'_n/(\log n)^{1+4\beta} \to \infty$ and $|\log(h''_n)|/\log\log n \to \infty$ as $n \to \infty$, we have*

$$\lim_{n\to\infty} \sup_{h'_n \leqslant h \leqslant h''_n} \big|H_{n,h,\beta}(f) - H(f)\big| = 0 \quad \textit{a.s.} \tag{6}$$

We note that the main problem in using entropy estimates such as (4) is to choose properly the smoothing parameter h_n. The uniform in bandwidth consistency result given in (6) shows that any choice of h between h'_n and h''_n ensures the consistency of $H_{n,h,\beta}(f)$. Now, we shall establish another result in a similar direction for a class of compactly supported densities. We need the following conditions:

(F.2) f has a compact support say $\mathbb{I}$, and there exists a constant $0 < \mathfrak{M} < \infty$ such as

$$\sup_{\mathbf{x}\in\mathbb{R}^d} \left|\frac{\partial^s f(\mathbf{x})}{\partial x_1^{j_1} \ldots \partial x_d^{j_d}}\right| \leqslant \mathfrak{M}, \quad j_1 + \cdots + j_d = s.$$

(K.5) K is of order s, i.e., for some constant $\mathfrak{S} \neq 0$,

$$\int_{\mathbb{R}^d} t_1^{j_1} \ldots t_d^{j_d} K(\mathbf{t})\,\mathrm{d}\mathbf{t} = 0, \quad j_1, \ldots, j_d \geqslant 0,\ j_1 + \cdots + j_d = 1, \ldots, s-1,$$

$$\int_{\mathbb{R}^d} \big|t_1^{j_1} \ldots t_d^{j_d}\big| K(\mathbf{t})\,\mathrm{d}\mathbf{t} = \mathfrak{S}, \quad j_1, \ldots, j_d \geqslant 0,\ j_1 + \cdots + j_d = s.$$

Under the condition (F.2), the differential entropy of f may be written as follows

$$H(f) = -\int_{\mathbb{I}} f(\mathbf{x}) \log\big(f(\mathbf{x})\big)\, d\mathbf{x}.$$

Theorem 2.3. *Let K satisfy (K.1-2-3-4-5), and let f fulfill (F.1-2). Then for any $\beta > 0$, and for each pair of sequences $0 < h'_n < h''_n \leqslant 1$ with $h''_n \to 0$ and $nh'_n/(\log n)^{1+4\beta} \to \infty$ as $n \to \infty$, we have*

$$\limsup_{n\to\infty} \sup_{h'_n \leqslant h \leqslant h''_n} \frac{\sqrt{nh}|H_{n,h,\beta}(f) - H(f)|}{\sqrt{\{\log n\}^{4\beta}(\log(1/h) \vee \log\log n)}} \leqslant \zeta(\mathbb{I}) \quad a.s.,$$

where

$$\zeta(\mathbb{I}) = \sup_{\mathbf{x}\in\mathbb{I}} \left\{ f(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{u})\, d\mathbf{u} \right\}^{1/2}.$$

Remark 1. Theorem 2.3 leads to the construction of asymptotic 100% certainty interval for the true entropy $H(f)$, using the techniques developed in [5]. We give in what follows, the idea how to construct this interval. Throughout, we let $h \in [h'_n, h''_n]$, where h'_n and h''_n are as in Theorem 2.3. We infer from Theorem 2.3 that, for suitably chosen data-dependent functions $L_n = L_n(X_1, \ldots, X_n) > 0$, for each $0 < \varepsilon < 1$ and for any $\beta > 0$, we have, as $n \to \infty$,

$$\mathbb{P}\left(\frac{1}{L_n} \big| H_{n,h,\beta}(f) - H(f) \big| \geqslant 1 + \varepsilon \right) \to 0. \tag{7}$$

Assuming the validity of the statement (7), we obtain asymptotic certainty interval for $H(f)$ in the following sense. For each $0 < \varepsilon < 1$, we have, as $n \to \infty$,

$$\mathbb{P}\big(H(f) \in \big[H_{n,h,\beta}(f) - (1+\varepsilon)L_n,\, H_{n,h,\beta}(f) + (1+\varepsilon)L_n\big]\big) \to 1. \tag{8}$$

Whenever (8) holds for each $0 < \varepsilon < 1$, we will say that the interval

$$\big[H_{n,h,\beta}(f) - L_n,\, H_{n,h,\beta}(f) + L_n\big]$$

provides asymptotic 100% certainty interval for $H(f)$.

To construct L_n we proceed as follows. Assume that there exists a sequence $\{\mathbb{I}_n\}_{n\geqslant 1}$ of strictly nondecreasing compact subsets of $\mathbb{I}$, such that $\bigcup_{n\geqslant 1} \mathbb{I}_n = \mathbb{I}$ (for the estimation of the support $\mathbb{I}$ we may refer to [8] and the references therein). Furthermore, suppose that there exists a sequence (possibly random) $\{\zeta_n(\mathbb{I}_n)\}, n = 1, 2, \ldots,$ converging to $\zeta(\mathbb{I})$ in the sense that

$$\mathbb{P}\left(\left| \frac{\zeta_n(\mathbb{I}_n)}{\zeta(\mathbb{I})} - 1 \right| \geqslant \varepsilon \right) \to 0 \quad \text{as } n \to \infty \text{ for each } \varepsilon > 0. \tag{9}$$

Observe that the statement (9) is satisfied when the choice $\zeta_n(\mathbb{I}_n) = \sup_{\mathbf{x}\in\mathbb{I}_n} \sqrt{f_{n,h}(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{u})\, d\mathbf{u}}$ is considered. Consequently, we may define the quantity L_n displayed in the statement (7) by

$$L_n = \sqrt{\frac{\{\log n\}^{4\beta}(\log(1/h) \vee \log\log n)}{nh}} \times \zeta_n(\mathbb{I}_n).$$

Remark 2. Giné and Mason [11] establish uniform in bandwidth consistency for the one-live-out entropy estimator, which is defined by

$$\widehat{H}_{n,h_n} = -\frac{1}{n} \sum_{i=1}^{n} \log\big\{ f_{n,h_n,-i}(X_i) \big\},$$

where

$$f_{n,h_n,-i}(X_i) = 1/\big((n-1)h_n\big) \sum_{1\leqslant j\neq i\leqslant n} K\big((X_i - X_j)/h_n\big).$$

Their results hold subject to the condition that the density f is bounded away from 0 on its support.

3. Sketch of the proofs

Proof of Theorem 2.1. The proof is based essentially on the following decomposition:

$$H_{n,h_n,\beta}(f) - \widehat{\mathbb{E}}H_{n,h_n,\beta}(f) = -\int_{A_{n,\beta}} \big\{\log f_{n,h_n}(\mathbf{x}) - \log \mathbb{E} f_{n,h_n}(\mathbf{x})\big\}\mathbb{E} f_{n,h_n}(\mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{A_{n,\beta}} \big\{f_{n,h_n}(\mathbf{x}) - \mathbb{E} f_{n,h_n}(\mathbf{x})\big\}\log f_{n,h_n}(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

and the following fact due to [10]:

Fact 2. *Under the conditions of Theorem* 2.1*, we have for any* $c > 0$*, with probability* 1*,*

$$\limsup_{n\to\infty} \sup_{cn^{-1}\log n \leqslant h \leqslant 1} \frac{\sqrt{nh}\,\|f_{n,h} - \mathbb{E} f_{n,h}\|_\infty}{\sqrt{\log(1/h) \vee \log\log n}} = \Sigma(c). \qquad \square$$

Proof of Corollary 2.2. As a first step, we obtain the following evaluation of the bias term:

$$\big\|\mathbb{E} f_{n,h}(\mathbf{x}) - f(\mathbf{x})\big\|_\infty = O\big(\big(h_n''\big)^{1/d}\big).$$

This fact combined with Theorem 2.1, completes our proof. □

Proof of Theorem 2.3. Under the conditions of Theorem 2.3, we have

$$\big\|\mathbb{E} f_{n,h}(\mathbf{x}) - f(\mathbf{x})\big\|_\infty = O\big(\big(h_n''\big)^{s/d}\big).$$

This in combination with Theorem 2.1, completes our proof. □

Acknowledgements

We would like to thank the referees for their constructive criticism and helpful comments. We are also grateful to Professor Paul Deheuvels for his helpful discussions and suggestions leading to improvement of this Note.

References

[1] H. Akaike, An approximation to the density function, Ann. Inst. Statist. Math., Tokyo 6 (1954) 127–132.
[2] J. Beirlant, E.J. Dudewicz, L. Györfi, E.C. van der Meulen, Nonparametric entropy estimation: An overview, Int. J. Math. Stat. Sci. 6 (1) (1997) 17–39.
[3] D. Bosq, J.-P. Lecoutre, Théorie de l'estimation fonctionnelle. Économie et Statistiques Avancèes, Economica, Paris, 1987.
[4] P. Deheuvels, Uniform limit laws for kernel density estimators on possibly unbounded intervals, in: Recent Advances in Reliability Theory, Bordeaux, 2000, in: Stat. Ind. Technol., Birkhäuser Boston, Boston, MA, 2000, pp. 477–492.
[5] P. Deheuvels, D.M. Mason, General asymptotic confidence bands based on kernel-type function estimators, Stat. Inference Stoch. Process. 7 (3) (2004) 225–277.
[6] L. Devroye, L. Györfi, Nonparametric Density Estimation: The L_1 View, Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics, John Wiley & Sons Inc., New York, 1985.
[7] L. Devroye, G. Lugosi, Combinatorial Methods in Density Estimation, Springer Series in Statistics, Springer-Verlag, New York, 2001.
[8] L. Devroye, G.L. Wise, Detection of abnormal behavior via nonparametric estimation of the support, SIAM J. Appl. Math. 38 (3) (1980) 480–488.
[9] U. Einmahl, D.M. Mason, An empirical process approach to the uniform consistency of kernel-type function estimators, J. Theoret. Probab. 13 (1) (2000) 1–37.
[10] U. Einmahl, D.M. Mason, Uniform in bandwidth consistency of kernel-type function estimators, Ann. Statist. 33 (3) (2005) 1380–1403.
[11] E. Giné, D.M. Mason, Uniform in bandwidth estimation of integral functionals of the density function, Scand. J. Statist. 35 (4) (2008) 739–761.
[12] L. Györfi, E.C. van der Meulen, On the nonparametric estimation of the entropy functional, in: Nonparametric Functional Estimation and Related Topics, Spetses, 1990, in: NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 335, Kluwer Acad. Publ., Dordrecht, 1991, pp. 81–95.
[13] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Statist. 33 (1962) 1065–1076.
[14] B.L.S. Prakasa Rao, Nonparametric functional estimation, in: Probability and Mathematical Statistics, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983.
[15] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Statist. 27 (1956) 832–837.
[16] C.E. Shannon, A mathematical theory of communication, Bell System Tech. J. 27 (1948) 379–423, 623–656.