Advanced Mathematical Models & Applications
Vol.7, No.3, 2022, pp.333-338

ON CONVERGENCE RATES OF NONPARAMETRIC DENSITY FUNCTIONALS ESTIMATES*

Ibrahim A. Ahmad[1], **A.R. Mugdadi**[2†]

[1]Department of Statistics, Oklahoma State University, Stillwater, OK 74078, USA
[2]Department of Mathematics and Natural Sciences, Higher Colleges of Technology, Abu Dhabi, UAE

Abstract. Kernel density estimate is considered as the main nonparametric methods to estimate the probability density function. In this investigation we study the rate of convergence for two measures of deviation between the probability density function and the kernel probability density estimate. Rate of almost sure convergence are given. The results are extendable in straight forward way to multivariate case.

Keywords: density estimation, kernel method, strong consistency, rates of convergence.
AMS Subject Classification: 60F12, 60F05, 62G30.
†**Corresponding author:** A.R. Mugdadi, Department of Mathematics and Natural Sciences, Higher Colleges of Technology, Abu Dhabi, UAE, e-mail: *amugdadi@hct.ac.ae*
Received: 22 June 2022; Revised: 26 July 2022; Accepted: 26 August 2022; Published: 15 December 2022.

1 Introduction

Kernel density estimation is one of the classical topics in nonparametric statistics, starting from the work of Rosenblatt (1956) and Parzen (1962), nonparametric methods for density estimation were developed and applied to many areas in statistical inference. Large sample properties, such as consistency and asymptotic distributions are interesting to justify the use of these methods. One criterion for strong consistency is to measure the deviation of the estimate $\hat{f}(x)$ from the true density $f(x)$. Bickel and Rosenblatt (1973) proposed the measure

$$J_{1n}^2 = \int [\hat{f}(x) - f(x)]^2 dx. \tag{1}$$

An application of density estimation is to estimate the functional $\theta = \int f^2(x)dx$. The functional θ is important in many nonparametric inferential problems as it appears as the dominant term in many relative efficiencies expressions of rank statistics. An estimate of θ may be $\hat{\theta} = \int \hat{f}^2(x)dx$ or $\hat{\hat{\theta}} = \int \hat{f}(x)dF_n(x)$, where $F_n(x)$ denotes the empirical distribution function. Here we shall be concerned with the functional

$$J_{2n} = |\int \hat{f}^2(x)dx - \int f^2(x)dx|, \tag{2}$$

and

$$J_{2n}^* = |\int \hat{f}(x)dF_n(x) - \int f^2(x)dx|. \tag{3}$$

*How to cite (APA)**: Ahmad, I.A., Mugdadi, A.R. (2022). On convergence rates of nonparametric density functionals estimates. *Advanced Mathematical Models & Applications*, *7*(3), 333-338.

In the above and all what follows, whenever no limits of an integration are given it is taken over $(-\infty, \infty)$.

The kernel method is one of the most successful techniques for density estimation which can be described briefly as:

Let $X_1, ..., X_n$ be a random sample from $f(x)$. Let k be a known probability density function satisfying the following conditions: k is right continuous, $sup_u k(u) < \infty$, and $|u|k(|u|) \to 0$ as $|u| \to \infty$. Furthermore, let $\{a_n\}$ be a sequence of real numbers such that $a_n \to 0$ as $n \to \infty$. The kernel estimate of $f(x)$ using $k(u)$ is given by:

$$\begin{aligned} \hat{f}(x) &= \frac{1}{a_n} \int k(\frac{x-y}{a_n}) dF_n(y) \\ &= \frac{1}{a_n} \sum_{j=1}^{n} k(\frac{x-X_j}{a_n}). \end{aligned} \tag{4}$$

Kernel density estimate has application in Engineering, Agriculture, Computer Science and other fields, for example Parente et al. (2020) used Kernel density estimates for sepsis classification where Severe sepsis is a leading cause of intensive care unit (ICU) admission, Hewitt, et al (2022) used kernel density estimates satellite tag data wher they conditional distributions for quantitative comparison of pre- and post-exposure behavior. Taaffe, et al. (2021) implement kernel density estimation for the probability distribution of surgery duration.

We shall study the rates of convergence in strong consistency of J_{1n}^2, J_{2n}, and J_{2n}^* above. Devroye (1983) studied the L_1 distance as a measure of closeness between $\hat{f}(x)$ and $f(x)$. Kundu and Martinsek (1997) proposed procedures for bounding the L_1 distance. Ahmad and Mugdadi (2006) used the wieghted Hellinger distance as an error criterion to evalute the bandwidth for density estimation, later, Mugdadi and Anver (2016) used the weighted Hellinger distance as measure of error for the Multivariate kernel density estimate. Cheng (2019) considered under L_p-norm, the global property for the error density estimator with censored survival data. Rao (2010) studied the problem of estimation of density function by the method of delta sequences for functional data, Nadar (2010) studied the local convergence rate for the mean square error for the multivariate density estimation when the density function satisfies certain conditions. Karunamuni et al (2006) investigated the convergence properties of an adaptive kernel density estimation under some regularity conditions. recently Ahmad and Mugdadi (2020) obtianed the mean square error as a measure of error for for the L-estimator.

Ahmad (1976 a) showed that $J_{1n}^2 \to 0$ with probability one (w.p.1) under certain conditions while Ahmad (1976 b) showed that $J_{2n} \to 0$ and $J_{2n}^* \to 0$ w.p.1 as $n \to \infty$. The following result due to Kuelbs (1977) is important in the sequel. The proof of this Theorem rests on a result of Kuelbs (1978).

Theorem 1. *Assume that k is a function of bounded variation such that $\int |u|k(u)du < \infty$. Let f satisfy the condition*

$$sup_x |f(x)| + sup_{x \neq y} \frac{|f(x) - f(y)|}{|x-y|} < \infty. \tag{5}$$

Then

(i) $sup_x |\hat{f}(x) - E\hat{f}(x)| = O(\sqrt{\frac{log(log(n))}{n}} a_n^{-1})$,

(ii) $M = sup_n sup_x |\hat{f}(x) - E\hat{f}(x)| \sqrt{\frac{n}{log(log(n))}} a_n$ is a random variable such that for all $\beta > 0$, $E(e^{\beta M^2}) < \infty$,

(iii) For $a_n = n^{-1/4}$,

$$sup_x |\hat{f}(x) - f(x)| = O(n^{-1/4} \sqrt{log(log(n))}),$$

(iv) For every $\beta > 0$*, and if* $\sqrt{\frac{log(log(n))}{n}}a_n^{-1} \to 0$ *as* $n \to \infty$*,*

$$lim_{n\to\infty}E(e^{\beta sup_x|\hat{f}(x)-f(x)|^2}) = 1.$$

Remark 1. *If we allow stronger conditions, the rate in (iii) may be improved. In fact if we assume that* $f(x)$*,* $f'(x)$ *and* $f''(x)$ *are all uniformly bounded, if* k *is a function of bounded variation such that* $\int uk(u)du = 0$ *and* $\int u^2k(u) < \infty$ *then with* $a_n = n^{-1/6}$ *we have* $sup_x|\hat{f}(x) - f(x)| = O(n^{-1/3}(log(log(n)))^{1/2})$*, see Theorem D of Kuelbs (1977).*

In Section 2 we shall obtain strong consistency analogous results for J_{1n}^2*,* J_{2n}*, and* J_{2n}^**. It should be noted here that the proofs extend to the multivariate case without difficulty.*

2 Main results

Theorem 2. *Assume that the conditions of Theorem 1 are satisfied and that* $f(x)$ *is square integrable. Let* $J_{1n}^2 = \int[\hat{f}(x) - f(x)]^2dx$ *and* $L_{2n}^2 = \int[\hat{f}(x) - E\hat{f}(x)]^2dx$*. Then*

(i) $L_{1n} = O(\sqrt{\frac{log(log(n))}{na_n}})$.

(ii) If $M = sup_n L_{1n}\sqrt{\frac{log(log(n))}{na_n^2}}$*, then* $E(exp(\beta M^2)) < \infty$.

(iii) If $a_n = n^{\frac{-1}{4}}$*, then* $J_{1n} = O(n^{\frac{-1}{4}}\sqrt{log(log(n))})$.

(iv) For each $\beta > 0$*,* $\frac{log(log(n))}{na_n^2} \to 0$ *as* $n \to \infty$*,* $E(exp(\beta J_{1n}^2) \to 1$ *as* $n \to \infty$.

Proof. It follows from Theorem 1 of Ahmad (1976 a), that

$$L_{1n}^2 \leq 4\delta[sup_x|\hat{f}(x) - E\hat{f}(x)|]^2 \tag{6}$$

for some $\delta > 0$, thus $L_{1n} \leq Csup_x|\hat{f}(x) - E\hat{f}(x)|$. Hence (i) and (ii) follow from (i) and (iii) of Theorem 1. From relation (2.12) of Ahmad (1976 a) we have,

$$[\int(E\hat{f}(x) - f(x))^2dx]^{1/2} \leq \frac{1}{a_n}\int k(\frac{v}{a_n})[\int(f(x-v) - f(x))^2dx]^{1/2}dv. \tag{7}$$

Now,

$$\int[f(x-v) - f(x)]^2dx \leq sup_x|f(x-v) - f(x)|\int|f(x-v) - f(x)|dx.$$

Let $\delta > 0$ be any real number and set $N(\delta) = (-\infty, -\delta] \cup [\delta, \infty)$. Thus

$$\begin{aligned}\int[f(x-v) - f(x)]dx &\leq 2\delta sup_x|f(x-v) - f(x)| + \int_{N(\delta)}|f(x-v) - f(x)|dx \\ &\leq 2\delta sup_x|f(x-v) - f(x)| + \int_{-\delta}^{\delta}[f(x-v) - f(x)]dx \\ &\quad +2\int_{N(\delta)}f(x)dx \\ &\leq 4\delta sup_x|f(x-v) - f(x)| + 2\int_{N(\delta)}f(x)dx, \end{aligned} \tag{8}$$

since f is a probability density function, then for every $\epsilon > 0$ such that $\int_{N(\delta)}f(x) < \epsilon$, and since ϵ is arbitrary we have

$$\int[f(x-v) - f(x)]dx \leq 4\delta sup_x|f(x-v) - f(x)|.$$

Hence

$$
\begin{aligned}
(\int [E\hat{f}(x) - f(x)]^2 dx)^{1/2} &\leq \frac{C}{a_n}\int k(\frac{v}{a_n}) sup_x |f(x-v)f(x)| dxdv \\
&\leq \frac{C_1}{a_n},
\end{aligned} \tag{9}
$$

since $\int |u|k(u)du < \infty$ and $\frac{sup_x|f(x-v)-f(x)|}{|v|} < \infty$ for every $|v| > 0$. Hence (ii) and (iv) follow from (i) and (ii). This concludes the proof. □

Theorem 3, to follow, gives analogous results about J_{2n}, while Theorem 4 gives results corresponding the modified version $J_{2n}^* = |\int \hat{f}(x)dF_n(x) - \int f^2(x)dx|$.

Theorem 3. *Let $\theta = \int f^2(x)dx$ and $\hat{\theta} = \int \hat{f}^2(x)dx$ where $\hat{f}(x)$ is as given in (4), and assume that the conditions of Theorem 2 are satisfied. Then*

(i) $|\hat{\theta} - E\hat{\theta}| = O(\sqrt{\frac{log(log(n))}{n}}\frac{1}{a_n})$.

(ii) If $M = sup_n|\hat{\theta} - E\hat{\theta}|\sqrt{\frac{log(log(n))}{na_n^2}}$, then for any $\beta > 0$, $E(exp(\beta M^2)) < \infty$.

(iii) If $a_n = n^{\frac{-1}{4}}$, then $J_{1n} = O(n^{\frac{-1}{4}}\sqrt{log(log(n))})$.

(iv) If $\frac{loglog\, n}{na_n^2} \to 0$ as $n \to \infty$, then $E(exp(\beta J_{2n})) \to 1$ as $n \to \infty$.

Proof. Note that

$$
\begin{aligned}
|\hat{\theta} - E\hat{\theta}| &\leq \int |\hat{f}(x) - E\hat{f}(x)|\hat{f}(x)dx + \int |\hat{f}(x) - E\hat{f}(x)|E\hat{f}(x)dx \\
&\leq 2sup_x|\hat{f}(x) - E\hat{f}(x)|,
\end{aligned} \tag{10}
$$

since $\int E\hat{f}(x)dx = \int \hat{f}(x)dx = 1$. Thus (i) and (ii) follow from (i) and (ii) of Theorem 1. Next,

$$
\begin{aligned}
J_{2n} &= |\hat{\theta} - \theta| \\
&\leq \int |\hat{f}(x) - f(x)|\hat{f}(x)dx + \int |\hat{f}(x) - f(x)|f(x)dx \\
&\leq 2sup_x|\hat{f}(x) - f(x)|.
\end{aligned} \tag{11}
$$

Thus (iii) and (iv) follow from (iii) and (iv) of Theorem 1. □

Theorem 4. *Let $\hat{\hat{\theta}} = \int \hat{f}(x)dF_n(x)$ and $\tilde{\theta} = \int E\hat{f}(x)f(x)dx$, and assume that the conditions of Theorem 1 are satisfied. Then*

(i) $|\hat{\hat{\theta}} - \tilde{\theta}| = O(\sqrt{\frac{log(log(n))}{n}}\frac{1}{a_n})$.

(ii) If $M = sup_n|\hat{\hat{\theta}} - \tilde{\theta}|\sqrt{\frac{log(log(n))}{na_n^2}}$, then for any $\beta > 0$, $E(exp(\beta M^2)) < \infty$.

(iii) If $a_n = n^{\frac{-1}{4}}$, then $J_{2n}^* = |\hat{\hat{\theta}} - \tilde{\theta}| = O(n^{\frac{-1}{4}}\sqrt{log(log(n))})$.

(iv) If $\frac{log(log(n))}{na_n^2} \to 0$ as $n \to \infty$, then for every $\beta > 0$, $E(exp(\beta J_{2n}^*)) \to 1$ as $n \to \infty$.

Proof. Note that

$$
\begin{aligned}
|\hat{\hat{\theta}} - \tilde{\theta}| &\leq \int |\hat{f}(x) - E\hat{f}(x)|dF_n(x) + |\int E\hat{f}(x)d(F_n(x) - F(x))| \\
&\leq sup_x|\hat{f}(x) - E\hat{f}(x)| + \frac{C}{a_n}sup_x|F_n(x) - F(x)| \\
&\leq \frac{D}{a_n}sup_x|F_n(x) - F(x)|,
\end{aligned} \tag{12}
$$

where the second term in the second inequality is obtained by integration by parts and since (cf. Nadaraya (1965))

$$sup_x|\hat{f}(x) - E\hat{f}(x)| \leq \frac{D}{a_n} sup_x|F_n(x) - F(x)|, \tag{13}$$

the third inequality follows with some constant $D > 0$. Thus, (i) and (iIi) follows exactly as in the proof of Theorem 1 (cf. Kuelbs (1977)). Next (ii) and (iv) follow directly as in (ii) and (iv) of Theorem 1 (cf. Kuelbs (1978)) with the aid of the inequality

$$|\tilde{\theta} - \theta| \leq \int |E\hat{f}(x) - f(x)|dF(x) \leq sup_x|E\hat{f}(x) - f(x)|. \tag{14}$$

The Theorem is proved. □

Acknowledgment: The authors thank the managing editor and the referees for their remarks on the earlier version of this article which improve the last version of the paper.

References

Ahmad, I.A. (1976a). Strong convergence of a quadratic measure of the deviation of denisty function estimates. *Rep. Statist. Appl. Res. JUSE*, 25, 1-5.

Ahmad, I.A. (1976b). On asymptotoc properties of an estimate of a functional of probability density. *Scan. Actuarial J.*, 4, 176-181.

Ahmad, I.A., Mugdadi, A.R. (2006). Weighted Hellinger distance as an error criterion for bandwidth selection in kernel estimation. *Journal of Nonparametric Statistics*, *18*(2), 215-226.

Ahmad, I.A., Mugdadi, A.R. (2020). On the mean square error of smooth L-estimator. *Missouri of Journal of Mathematical Sciences,* 32,2, 121-127.

Bickel, P.J., Rosenblatt, M. (1973). On some global measures of the deviation of density function estimates. *Ann. Statist.*, 1, 1079-1095.

Cheng, F. (2019). The L_p consistency of error density estimator in censored linear regression. *Communications in Statistics - Theory and Methods*, *48*(7), 1579-1584.

Devroye, L. (1983). The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.*, 11, 896-904.

Hewitt, J., Gelfand, A., Quick, N., Cioffi, W., Southhall, B., DeRuiter, S., & Schick, R. (2022). Kernel density estimation of conditional distributions to detect responses in satellite tag data, *Animal Biotelemetry,* 10,1, Article 28.

Karunamuni, R., Sriram, T. & Wu, J. (2006). Rates of convergence of an adaptive kernel density estimator for finite mixture models. *Statistics and Probability Letters*, 76, 221-230.

Kuelbs, J. (1977). Some exponential moments with applications to density estimation, the empirical distribution function, and lacunary series. *Proceedings of the Dublin Conference on Vector Measures 1977*, Lecture Notes in Mathematics 644, Springer Verlag, New York.

Kuelbs, J. (1978). Some exponential moments of sums of independent random rariables. *Transactions of the American Mathematical Society*, 240, 145-162.

Kundu, S., Martinsek, A. (1997). Bounding the L_1 distance in nonparametric density estimation. *Ann. Inst. Stat. Math.,* 49, 57–78.

Mugdadi, A.R., Anver, H. (2016). The weighted hellinger distance in the multivariate kernel density estimation. *South African Statistical Journal,* 50,2,221-236

Nadar, M. (2010). Local convergence rate of mean squared error in density estimation. *Communications in Statistics - Theory and Methods*, 40, 176-185.

Nadaraya, E. (1965). On nonparametric estimation of density function and regression curves. *Theory Probab. Appl.*, 10, 186-190.

Parente, J., Chase, J., Moller, K. & Shaw, G. (2020). Kernel density estimates for sepsis classification. *Computer Methods and Programs in Biomedicine,* 188, Article 105295.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.,* 33, 1065-1076.

Rao, B.P. (2010). Nonparametric density estimation for functional data by delta sequences. *Brazilian Journal of Probability and Statistics*, 24, 3, 468-478.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a denisty function. *Ann. Math. Statist.,* 27, 832-837.

Taaffe, K., Pearce, B. & Ritchie, G. (2021). Using kernel density estimation to model surgical procedure duration. *International Transactions in Operational Research,* 28, 1, 401-418.