Statistics and Probability Letters 82 (2012) 1786–1791

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

An optimal *k*-nearest neighbor for density estimation

Yi-Hung Kung [a], Pei-Sheng Lin [a,b,*], Cheng-Hsiung Kao [b]

[a] *Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taiwan*

[b] *Department of Mathematics, National Chung Cheng University, Taiwan*

ARTICLE INFO

Article history:
Received 20 December 2011
Received in revised form 17 May 2012
Accepted 18 May 2012
Available online 4 June 2012

Keywords:
Density estimation
Multi-dimensional data
Nearest neighbor method

ABSTRACT

A *k*-nearest neighbor method, which has been widely applied in machine learning, is a useful tool to obtain statistical inference for an underlying distribution of multi-dimensional data. However, the knowledge on choosing an optimal order for the *k*-nearest neighbor is relatively little. This paper proposes an asymptotic distribution for the nearest neighbor statistic. Under some conditions, we find an optimal unbiased density estimate based on a linear combination of nearest neighbors, and it leads to an optimal choice for the order of the *k*-nearest neighbor.

1. Introduction

Density estimation has been proved to be a useful tool for implementation of exploratory data analysis. For example, to obtain statistical inference in a regression model, we may need to know the probability density function of a noise random variable, or whether the underlying distribution is skew. Since data are often considered as a random sample from an unknown probability function, a nonparametric approach could provide more flexibility in application of density estimation. Although a histogram estimate is the earliest concept of non-parametric density estimation, the maturer techniques are evolved with kernel density estimation (e.g. Parzen, 1962). Since the work by Parzen, various kernel density estimation methods have been proposed (e.g. Scott, 1992, Prakasa Rao, 1983, and Silverman, 1986). However, Silverman (1986) mentioned that applying the kernel density estimation to multivariate data may be difficult because of 'curse of dimensionality'.

On the other hand, the *k*-nearest neighbor method, which has been widely used in machine learning, provides a simple tool for density estimation in multi-dimensional spaces. Li (1984) suggested to use a cross-validation (CV) method to find suitable *k* for density estimation. He also showed that the CV method is consistent under some regularity conditions. Biau et al. (2011) extended Li's work by implementing the CV method to a weighted function of nearest neighbors with some simulation studies.

Nevertheless, the research work for probability properties of the *k*-nearest neighbor method is still relatively little (e.g. Mack and Rosenblatt, 1979; Hall et al., 2008). Due to lack of enough knowledge for the probability distribution of the nearest neighbor statistic, choosing an optimal order of the *k*-nearest neighbor is a difficult task. In Section 2, we first explore a limiting distribution for the nearest neighbor statistic in a multiple dimensional space. Then, under some conditions, we find an asymptotically optimal unbiased estimate based on a linear combination of nearest neighbors. We also conduct simulations to compare the proposed method with other approaches in Section 3. Some discussions are given in the final section.

* Corresponding author at: Division of Biostatistics and Bioinformatics, National Health Research Institutes, 35 Keyan Road, Zhunan, Miaoli 350, Taiwan.
E-mail addresses: pslin@nhri.org.tw, pslin@math.ccu.edu.tw (P.-S. Lin).

0167-7152/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2012.05.017

2. Probability density estimation

Let $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ be a sample of event locations in a R^d-space. Assume that the event location $\boldsymbol{s}$ follows a common distribution with density function $f(\boldsymbol{s})$. For a given $\boldsymbol{s}_0 \in R^d$, we define $D_i(\boldsymbol{s}_0) = \|\boldsymbol{s}_i - \boldsymbol{s}_0\|,\ i = 1, 2, \ldots, n$, to be a Euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_0$. Let $D_{(1)}(\boldsymbol{s}_0) \le D_{(2)}(\boldsymbol{s}_0) \le \cdots \le D_{(n)}(\boldsymbol{s}_0)$ be the ordered statistics of $D_1(\boldsymbol{s}_0), \ldots, D_n(\boldsymbol{s}_0)$. Define $U_k = nD^d_{(k)}$. We also define $B(\boldsymbol{s}_0, r)$ to be a d-dimensional ball centered at $\boldsymbol{s}_0$ with radius r, and $N(\boldsymbol{s}_0, r)$ and $V(\boldsymbol{s}_0, r)$ to be the number of observations and volume in $B(\boldsymbol{s}_0, r)$, respectively. We first derive the following theorem.

Theorem 1. *Assume that* (a) *the density function $f(\boldsymbol{s})$ is bounded and uniformly continuous, and* (b) $k = O(n^{1/d})$ *and* $k \to \infty$ *as* $n \to \infty$. *Then, the limiting distribution of U_k follows a Gamma distribution.*

Proof. Let $g_k(\cdot)$ denote the probability density function of $U_k(\boldsymbol{s}_0)$. Since $P\{U_k(\boldsymbol{s}_0) > t\} = P\{N(\boldsymbol{s}_0, \tau_n) \le k-1\}$, where $\tau_n = (t/n)^{1/d}$, we have

$$\int_t^\infty g_k(s)ds = \sum_{j=0}^{k-1} P\{N(\boldsymbol{s}_0, \tau_n) = j\}. \tag{1}$$

When the number of observations n is large, by assumptions of Theorem 1, we have

$$P\{N(\boldsymbol{s}_0, \tau_n) = j\} \approx \binom{n}{j} \{f(\boldsymbol{s}_0)V(\boldsymbol{s}_0, \tau_n)\}^j \{1 - f(\boldsymbol{s}_0)V(\boldsymbol{s}_0, \tau_n)\}^{n-j}.$$

Let $C_d = \pi^{d/2}/\Gamma(d/2+1)$ and $\lambda_0 = f(\boldsymbol{s}_0)C_d$. Since $V(\boldsymbol{s}_0, \tau_n) = C_d\tau_n^d$ (Wegman, 1990), the above approximation can be rewritten as

$$P\{N(\boldsymbol{s}_0, \tau_n) = j\} \approx \binom{n}{j} (\lambda_0 t/n)^j (1 - \lambda_0 t/n)^{n-j}. \tag{2}$$

It is easy to see that, as $n \to \infty$, $\binom{n}{j}(\lambda_0 t/n)^j \to (\lambda_0 t)^j/j!$ and $(1 - \lambda_0 t/n)^{n-j} \to \exp(-\lambda_0 t)$. So, taking limits on both sides of (2) gives

$$\lim_{n\to\infty} P\{N(\boldsymbol{s}_0, \tau_n) = j\} = (\lambda_0 t)^j \exp(-\lambda_0 t)/j!. \tag{3}$$

Plugging (3) into (1) thus gives

$$\lim_{n\to\infty} \int_t^\infty g_k(s)ds = \sum_{j=0}^{k-1} \frac{(\lambda_0 t)^j}{j!} \exp(-\lambda_0 t).$$

Taking derivatives with respect to t on both sides of the above equation gives $g_k(t) \to \lambda_0(\lambda_0 t)^{k-1}\exp(-\lambda_0 t)/(k-1)!$ as $n \to \infty$, which is the density function of a gamma distribution. □

From Theorem 1, we learn that $U_k(\boldsymbol{s}_0)$ converges in distribution to a Gamma random variable with parameters k and $f(\boldsymbol{s}_0)C_d$ as $n \to \infty$. Fig. 1 depicts how the finite sample distribution of U_k converges to the asymptotic Gamma distribution at $\boldsymbol{s}_0 = (0, 0)$. In Fig. 1, we generated the event locations randomly from a mixture normal population $0.5\boldsymbol{X} + 0.5\boldsymbol{Y}$ and a standard normal population $0.5\boldsymbol{X}$, respectively, in various sample sizes $n = 100, 500, 1000$. (Details about the distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be seen in Section 3.) We run 1000 replications in each simulation setting.

As can be seen from Fig. 1, the finite sample distribution of U_k seems quite close to the asymptotic distribution given in Theorem 1. However, for the mixture normal case, we find from the histogram of simulated U_k values that U_k may produce some bias. So, we consider adding a multiplier for U_k to reduce estimation bias. By Slutsky's theorem, U_k^{-1} has a limiting inverse gamma distribution. Thus, for k satisfying assumptions of Theorem 1, it follows from Lebesgue's dominated convergence theorem that

$$E\left\{U_k^{-1}(\boldsymbol{s}_0)\right\} \to f(\boldsymbol{s}_0)C_d/(k-1), \tag{4}$$

as $n \to \infty$. Define $U_k^* = (k-1)U_k^{-1}/C_d$. From (4), we have the following result.

Corollary 1. *Under the assumptions of Theorem 1, $U_k^*(\boldsymbol{s}_0)$ is an asymptotically unbiased estimate for $f(\boldsymbol{s}_0)$.*

To use a linear combination of $U^*_{k_1}(\boldsymbol{s}_0), \ldots, U^*_{k_m}(\boldsymbol{s}_0)$ to estimate $f(\boldsymbol{s}_0)$, an optimal method would be to find $\omega_1, \ldots, \omega_k$ such that $E\{(\omega_1 U^*_{k_1} + \cdots + \omega_m U^*_{k_m})(\boldsymbol{s}_0)\} = f(\boldsymbol{s}_0)$ and $\mathrm{var}\{(\omega_1 U^*_{k_1} + \cdots + \omega_m U^*_{k_m})(\boldsymbol{s}_0)\}$ is minimized. The interpolation based on observed events $\boldsymbol{s}_{k_1}, \ldots, \boldsymbol{s}_{k_m}$ for a density of $\boldsymbol{s}_0$ can be found by using Lagrange's multiplier formula. Let $\boldsymbol{\Gamma}$ be a covariance

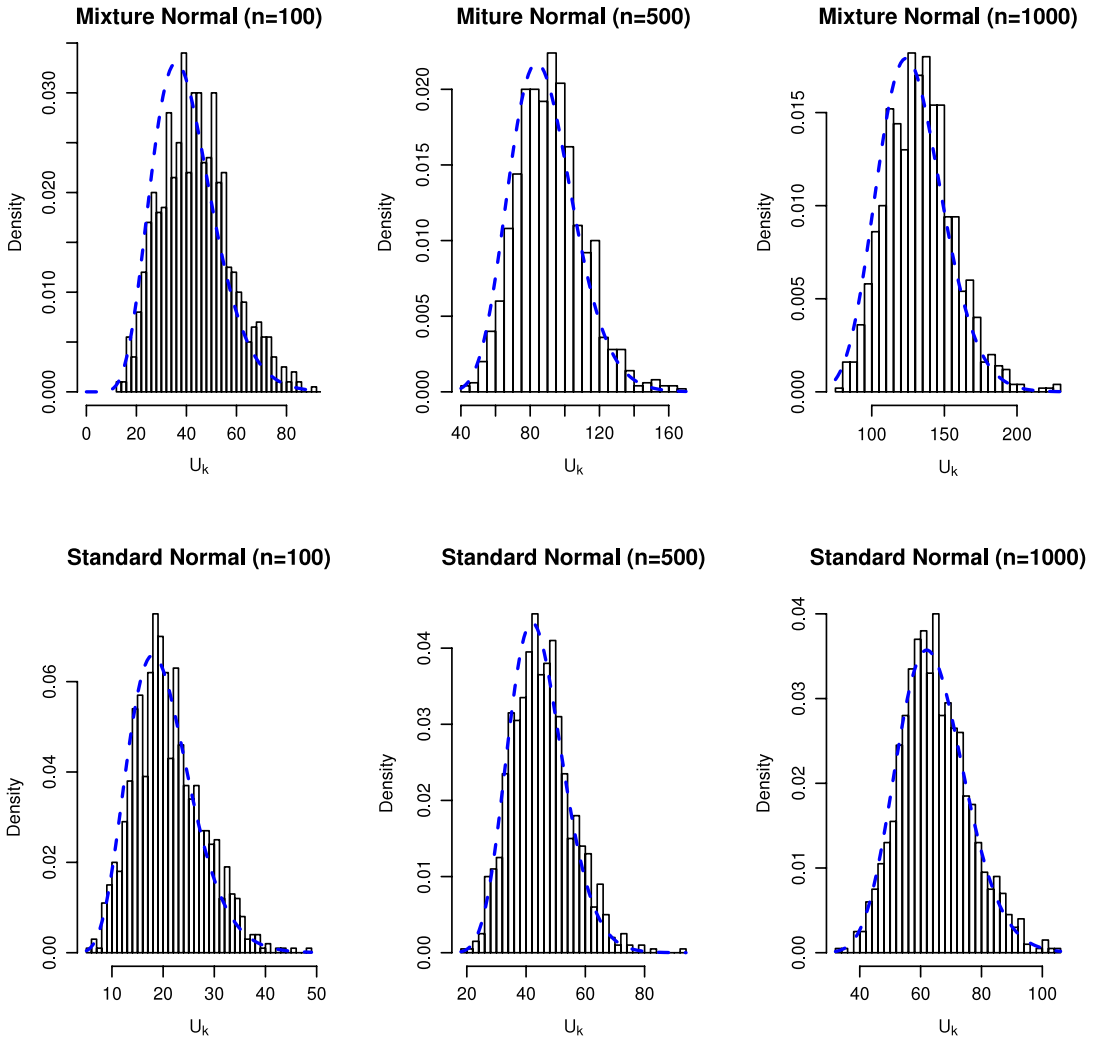

Fig. 1. A comparison of the finite sample and asymptotic distributions for the nearest neighbor density statistic U_k. The histograms are based on the simulated U_k values, and the dotted lines denote the Gamma density function based on Theorem 1.

matrix with the (i, j) element $\mathbf{\Gamma}_{i,j} = \mathrm{cov}(U_i^*, U_j^*)$ and $\mathbf{1}$ be a R^n-vector with all elements equal to one. It can be shown that, when

$$(\omega_1^*, \ldots, \omega_m^*)' = \frac{1}{\mathbf{1}'\mathbf{\Gamma}^{-1}\mathbf{1}}\mathbf{1}'\mathbf{\Gamma}^{-1}, \tag{5}$$

$\omega_1^* U_{k_1}^* + \cdots + \omega_m^* U_{k_m}^*$ is an optimal unbiased estimate based on a linear combination of $U_{k_1}^*, \ldots, U_{k_m}^*$ for $f(\boldsymbol{s}_0)$. Note that (5) is similar to the kriging model in spatial statistics. Next, we compute the covariance matrix $\mathbf{\Gamma}$ to simplify (5).

Theorem 2. *Under assumptions of Theorem 1, the limiting joint density of U_k and U_l, $l > k$, is*

$$f_{U_k,U_l}(t_1, t_2) = \frac{\lambda_0^2(\lambda_0 t_1)^{k-1}e^{-\lambda_0 t_1}}{(k-1)!}\frac{(\lambda_0 t_2 - \lambda_0 t_1)^{l-k-1}e^{-\lambda_0(t_2-t_1)}}{(l-k-1)!},$$

where $t_2 \geq t_1$, $t_1, t_2 \in (0, \infty)$.

Proof. We first compute $P\{U_k(\boldsymbol{s}_0) > t_1, U_l(\boldsymbol{s}_0) > t_2\}$. Let $\tau_{n_1} = (t_1/n)^{1/d}$ and $\tau_{n_2} = (t_2/n)^{1/2}$. Using an argument similar to the proof of Theorem 1, we have

$$\begin{aligned} &P\{N(\boldsymbol{s}_0, \tau_{n_1}) = j_1,\ N(\boldsymbol{s}_0, \tau_{n_2}) = j_2\} \\ &= \frac{n!}{j_1!(j_2-j_1)!(n-j_2)!}\left(\frac{\lambda_0 t_1}{n}\right)^{j_1}\left\{\lambda_0\left(\frac{t_2}{n} - \frac{t_1}{n}\right)\right\}^{j_2-j_1}\left(1 - \frac{\lambda_0 t_2}{n}\right)^{n-j_2}, \end{aligned} \tag{6}$$

which has a limit of $(\lambda_0 t_1)^{j_1}\lambda_0^{j_2-j_1}\exp(-\lambda_0 t_2)/j_1!(j_2-j_1)!$ as $n\to\infty$. Note that $P\{U_k(\boldsymbol{s}_0) > t_1, U_l(\boldsymbol{s}_0) > t_2\}$ is equal to a double summation of $P\{N(\boldsymbol{s}_0, \tau_{n_1}) = j_1, N(\boldsymbol{s}_0, \tau_{n_2}) = j_2\}$ over $j_1 = 0, \ldots, k-1$ and $j_2 = j_1, \ldots, l-1$. By Fubini's theorem and some simplifications for (6), we have

$$\lim_{n\to\infty} P\{U_k(\boldsymbol{s}_0) > t_1, U_l(\boldsymbol{s}_0) > t_2\} = \frac{t_1^{k-1}}{(k-1)!}\frac{(t_2-t_1)^{l-k-1}}{(l-k-1)!}\exp(-t_2).$$

The desired result then follows from taking derivatives of the above equality with respect to t_1 and t_2. □

We then calculate the covariance of U_k^* and U_l^* by using Theorem 2. Since $E(U_k^{-1}U_l^{-1}) \to \int_0^\infty\int_{t_1}^\infty (t_1t_2)^{-1}f_{U_k,U_l}(t_1,t_2)dt_2\,dt_1$ by Lebesgue's dominated theorem, after some changes of variables, we get

$$E\left(U_k^{-1}U_l^{-1}\right) \to \frac{\lambda_0^2}{k-1}\int_0^\infty\int_0^\infty \frac{1}{v+t}\frac{t^{k-2}e^{-t}}{(k-2)!}\frac{v^{l-k-1}e^{-v}}{(l-k-1)!}dv\,dt, \tag{7}$$

for k and l satisfying assumptions of Theorem 1. Let $T\sim$ Gamma$(k-1, 1)$ and $V\sim$ Gamma$(l-k, 1)$ be two independent Gamma random variables. Then, the joint density function within the integration of (7) can be regarded as a product of the densities of V and T. Since $V+T$ follows a Gamma distribution Gamma$(l-1, 1)$, to compute the integration of (7) is equal to computing the expected value of an inverse Gamma random variable with parameters $l-1$ and 1. So, $E\left(U_k^{-1}U_l^{-1}\right) \to \lambda_0^2/(k-1)(l-2)$ for $l>2$. Combining this result with Theorem 1 gives the following result.

Corollary 2. *Under the assumptions of Theorem 1, we have* $\mathrm{var}(U_k^*) = f^2(\boldsymbol{s}_0)/(k-2)$, $k>2$ *and* $\mathrm{cov}(U_k^*, U_l^*) = f^2(\boldsymbol{s}_0)/(l-2)$ *for* $l>2$ *and* $l>k$.

Suppose that we have a sequence of variables $U_{k_1}^*, \ldots, U_{k_m}^*$ with $k_1 < \cdots < k_m$ and k_j satisfying assumptions of Theorem 1. Plugging the covariance matrix $\boldsymbol{\Gamma}$ from Corollary 2 into (5), we have an interesting result from matrix algebra (Schott, 1997) that $\boldsymbol{\omega}^* = (0, \ldots, 0, 1)'$. That is, for $k_1 < \cdots < k_m$ satisfying assumptions of Theorem 1,

$$\omega_1^* U_{k_1}^*(\boldsymbol{s}_0) + \cdots + \omega_m^* U_{k_m}^*(\boldsymbol{s}_0) \equiv U_{k_m}^*(\boldsymbol{s}_0). \tag{8}$$

Corollary 3 is thus an immediate result from (8).

Corollary 3. *Let* $k_1 < \cdots < k_m$. *Under the assumptions of Theorem 1, an optimal linear unbiased estimate based on* $U_{k_1}^*, \ldots, U_{k_m}^*$ *is* $U_{k_m}^*$.

Furthermore, from Theorem 1 and Chebyshev's inequality, we have

$$P\left\{|k^{-1}U_k(\boldsymbol{s}_0) - \{C_d f(\boldsymbol{s}_0)\}^{-1}| < \epsilon\right\} \geq 1 - \{k\epsilon^2 C_d^2 f^2(\boldsymbol{s}_0)\}^{-1},$$

which approaches 1 as $k\to\infty$. This implies that

$$k^{-1}U_k(\boldsymbol{s}_0) \to \{C_d f(\boldsymbol{s}_0)\}^{-1} \quad \text{in probability,} \tag{9}$$

as n and k go to infinity. If we take $k = n^\alpha$, $\alpha\in(0, d^{-1})$, in (9), then $kU_k^*/(k-1)$ is a consistent estimator for $f(\boldsymbol{s}_0)$.

Corollary 4. *Under the condition of Theorem 1,* $n^\alpha U_{n^\alpha}^*/(n^\alpha-1)$, $\alpha\in(0, d^{-1})$, *is a consistent estimator for* $f(\boldsymbol{s}_0)$ *provided that* $f(\boldsymbol{s}_0)\neq 0$.

3. Simulation

In this section, a simple simulation study is conducted to compare the performance of the proposed method and the other k-nearest neighbor statistics by Mack and Rosenblatt (1979) and Biau et al. (2011). For convenience, we call the proposed method method NN_1 with density estimate $\hat{f}_1(\boldsymbol{s}_0) = (k_1-1)/\{nC_dD_{(k_1)}^d(\boldsymbol{s}_0)\}$ $(\equiv\sum_{j=2}^{k_1}\omega_j(j-1)/\{nC_dD_{(j)}^d(\boldsymbol{s}_0)\})$. Also, we refer to the work by Mack and Rosenblatt as method NN_2 with density estimate $\hat{f}_2(\boldsymbol{s}_0) = k_2/\{nC_dD_{(k_2)}^d(\boldsymbol{s}_0)\}$, and the work by Biau et al. as method NN_3 with density estimate $\hat{f}_3(\boldsymbol{s}_0) = \{k_3(1+k_3)\}/\{2nC_d\sum_{j=1}^{k_3}D_{(j)}^d(\boldsymbol{s}_0)\}$.

To generate event locations, let $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote bivariate normal random vectors with $\boldsymbol{X}\sim N_2\{(0,0)', \boldsymbol{I}_2\}$, where $\boldsymbol{I}_2$ is an identity matrix of size 2, and $\boldsymbol{Y}\sim N_2\{(4,0)', \boldsymbol{\Sigma}\}$, where $\boldsymbol{\Sigma}$ is a 2×2 matrix with the (i,j) element equal to 2 if $i=j$, and 1 otherwise. We then randomly simulated event locations $\boldsymbol{s}_i = (r_i, c_i)$, $i=1,\ldots,n$, from $\boldsymbol{X}$ or a mixture bivariate normal distribution $0.5\boldsymbol{X}+0.5\boldsymbol{Y}$. Based on the simulated event locations, we predicted density values on a $m\times m$ grid in a $(-3,3)\times(-3,3)$ region. In the simulation, we set the number of event locations $n = 50$, 100, 500, or 1000, the number of prediction points $m^2 = 900$ or 2500. For each simulation setting, 500 replicates were run.

Table 1
Under bivariate normal distributions, estimation bias and mean squared errors of methods NN_1 ($k_1 = Mn^{1/2}$), NN_2 ($k_2 = Mn^{2/3}$) and NN_3 (the CV method). Bold values denote the best cases. The listed biases and MSEs are 10^3 and 10^5 times the true values, respectively.

M	NN_1			NN_2			NN_3
	1.5	2.0	2.5	1.0	1.5	2.0	
	$(n, m) = (50, 30)$						
Bias	6.56	**6.33**	6.91	**6.54**	7.33	8.17	6.22
MSE	6.93	**6.03**	7.16	**6.33**	8.96	13.8	8.02
	$(n, m) = (50, 50)$						
Bias	6.73	**6.44**	6.96	**6.51**	7.39	8.26	6.24
MSE	7.67	**6.15**	7.06	**6.37**	8.88	14.2	8.46
	$(n, m) = (100, 30)$						
Bias	5.34	**5.30**	5.89	**5.67**	6.56	7.43	4.97
MSE	4.64	**4.37**	4.93	**4.59**	6.37	9.41	4.35
	$(n, m) = (100, 50)$						
Bias	**5.31**	5.46	5.93	**5.70**	6.62	7.51	5.26
MSE	4.56	**4.32**	5.07	**4.68**	6.53	9.69	4.72
	$(n, m) = (500, 50)$						
Bias	3.42	**3.41**	3.86	**4.06**	4.91	5.70	3.15
MSE	1.86	**1.82**	2.03	**2.21**	3.17	4.43	1.62
	$(n, m) = (1000, 50)$						
Bias	2.81	**2.80**	3.21	**3.52**	4.30	5.02	2.61
MSE	1.26	**1.21**	1.40	**1.63**	2.39	3.29	1.15

Note for method NN_1, an optimal order k_1 is $O(n^{1/d})$, while an optimal order for method NN_2 is $O\{n^{4/(d+4)}\}$. So, in the simulation, we chose $k_1 = Mn^{1/d}$ for method NN_1, and $k_2 = Mn^{d/(d+4)}$ for method NN_2, where $M = 1.0, 1.5, 2.0, 2.5$. On the other hand, we choose k_3 for method NN_3 based on the cross-validation criterion

$$\hat{k}_3 = \arg\min_k \sum_{i=1}^{n} \{\hat{f}_{-i}(\boldsymbol{s}_i) - f(\boldsymbol{s}_i)\}^2,$$

where $\hat{f}_{-i}(\boldsymbol{s}_i)$ is a leave-one-out k-nearest neighbor estimator of $f(\boldsymbol{s}_i)$ without using the sample of event location $\boldsymbol{s}_i$. To examine the performance of methods NN_1, NN_2, and NN_3, we compute the average of estimation bias by $\sum_{l=1}^{500}\sum_{i=1}^{m^2} |\hat{f}_j^{(l)}(\boldsymbol{s}_i) - f(\boldsymbol{s}_i)|/rm^2$ and the mean squared error by $\sum_{j=1}^{500}\sum_{i=1}^{m^2}\{\hat{f}_j^{(l)}(\boldsymbol{s}_i) - f(\boldsymbol{s}_i)\}^2/rm^2$, for each $j = 1, 2$, and 3.

Tables 1 and 2 show some of the simulation results for density estimates $\hat{f}_1, \hat{f}_2$, and $\hat{f}_3$ in the mixture normal (i.e., $0.5\boldsymbol{X} + 0.5\boldsymbol{Y}$) and standard normal (i.e., $\boldsymbol{X}$) situations, respectively. We used bold numbers to mark the best performance in each simulation setting. Overall, we find that the proposed method has the best performance at $M = 2.5$ under the standard normal situation, and $M = 2$ under the mixture normal situation. Nevertheless, even for the standard normal distribution, the performance of method NN_1 at $M = 2$ is still comparable with that at $M = 2.5$.

As can also be seen from Tables 1 and 2, when the sample size is moderate ($n = 50$ or 100), methods NN_1 and NN_2 have pretty close performance at their best cases. However, for large sample size n ($n = 500$ or 1000), method NN_1 seems to improve its prediction accuracy (i.e., small MSE) faster than method NN_2. In contrast, compared with method NN_3, method NN_1 has better performance in the moderate sample size. But for the mixture normal distribution, the convergence rate of method NN_3 looks better than method NN_1 as n increases, although method NN_1 also shows a competing convergence rate with method NN_3 under the standard normal distribution.

4. Conclusion and future research

In this paper, we find that the asymptotically optimal linear combination of nearest neighbors for density estimation is just the last term of the linear combination. A simple simulation study is also conducted to evaluate the performance of the proposed method in a finite sample. From the simulation study, we suggest to take $k = 2n^{1/d}$ for the order of the k-nearest neighbor estimate.

We note that the proposed order $O(n^{1/d})$ is smaller than the order $O\{n^{4/(d+4)}\}$ given by Mack and Rosenblatt (1979), which may present the issue that the convergency rate of the proposed method is slower than those of the Mack and Rosenblatt method and the CV method. However, we find from the simulation study that the proposed estimate could be better to distinguish two modes of a mixture normal distribution, especially when the sample size is small or moderate. Nevertheless, a more detailed computation for the convergency rate could provide a better idea for the choice of an optimal order.

Table 2
Under standard normal distributions, estimation bias and mean squared errors of methods NN_1 ($k_1 = Mn^{1/2}$), NN_2 ($k_2 = Mn^{2/3}$) and NN_3 (the CV method). Bold values denote the best cases. The listed biases and MSEs are 10^3 and 10^5 times the true values, respectively.

M	NN_1			NN_2			NN_3
	1.5	2.0	2.5	1.0	1.5	2.0	
	$(n, m) = (50, 30)$						
Bias	13.6	13.0	**12.7**	**12.7**	13.6	14.8	11.9
MSE	37.4	24.5	**24.1**	24.5	**24.3**	30.1	18.7
	$(n, m) = (50, 50)$						
Bias	14.0	**13.0**	13.6	**13.0**	13.8	15.1	12.2
MSE	39.7	25.9	**25.2**	**25.9**	26.0	31.9	21.2
	$(n, m) = (100, 30)$						
Bias	**10.6**	10.8	11.2	**11.0**	12.0	13.2	10.5
MSE	20.5	17.6	**17.1**	**17.2**	18.9	22.9	14.6
	$(n, m) = (100, 50)$						
Bias	10.9	11.0	**10.9**	**11.1**	12.1	13.3	10.6
MSE	21.9	18.0	**17.1**	**17.8**	18.9	23.1	15.0
	$(n, m) = (500, 50)$						
Bias	**6.69**	6.89	7.11	**7.49**	8.68	9.85	7.28
MSE	8.41	7.21	**7.02**	**7.26**	9.03	11.6	6.85
	$(n, m) = (1000, 50)$						
Bias	**5.62**	5.81	5.89	**6.49**	7.53	8.59	6.48
MSE	6.16	5.36	**5.22**	**5.48**	6.85	8.77	5.26

Acknowledgments

The author would like to thank the editor and an anonymous referee for the constructive comments toward improving the paper.

References

Biau, G., Chazal, F., Cohen-Steiner, D., Devroye, L., Rodríguez, C., 2011. A weighted k-nearest neighbor density estimate for geometric inference. Electronic Journal of Statistics 5, 204–237.
Hall, P., Park, B.U., Samworth, R.J., 2008. Choice of neighbor order in nearest-neighbor classification. The Annals of Statistics 36, 2135–2152.
Li, K.C., 1984. Consistency for cross-validated nearest estimates in nonparametric regression. The Annals of Statistics 12, 230–240.
Mack, Y.P., Rosenblatt, M., 1979. Multivariate k-nearest neighbor density estimates. Journal of Multivariate Analysis 9, 1–15.
Parzen, E., 1962. On estimation of a probability density function and mode. The Annals of Mathematical Statistics 33, 1065–1076.
Prakasa Rao, B.L.S., 1983. Nonparametric Functional Estimation. Academic Press, Orlando.
Schott, J.R., 1997. Matrix Analysis for Statistics. Wiley, New York.
Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, New York.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.
Wegman, E.J., 1990. Hyperdimensional data analysis using parallel coordinates. Journal of the American Statistical Association 85, 664–675.