SOME REMARKS ON OPTIMAL CHOICE OF k OF k-NEAREST NEIGHBOUR DENSITY ESTIMATION

By

M. VAJTA, Jr. and J. FRITZ*

Department of Automation, Technical University Budapest,

(Received November 8, 1974)

Presented by Prof. Dr. F. CSÁKI

Introduction

The investigations of different estimation methods on functions have become a common problem in the theory of recognizing and learning algorithms. These methods can be classified as parametrical and nonparametrical methods. In the case of parametrical-type methods the mathematical form of the density function $f(x)$ is supposed to be known (therefore the distribution is uniform, normal, exponential). During the training the typical parameters (e.g. expected value, variance) are estimated and a possible approximation $f(x)$ of the density function is produced, optimum by a goodness criterion.

There is generally no a priori, or very little information concerning the distribution of the obtained samples. The nonparametrical density estimation methods can be used in these cases. The main idea of these methods is: the value of a continuous density function at a point is estimated, proportional to the number of samples being in a small interval around the point [1, 8, 16, 21, 24].

One of the most generally used nonparametrical density estimation method is that developed by PARZEN. PARZEN introduced a so-called kernel function, by which $f(x)$ (approximating $\hat{f}(x)$ can be determined at a point using the following formula [1,8]:

$$\hat{f}(x) = \frac{1}{h(N)} \cdot \sum_{i=1}^{N} K\left(\frac{x - x_i}{h(N)}\right) \tag{1}$$

where N is the number of the samples, $h(N)$ is the scale distribution, $K(.)$ is the kernel function.

Using certain conditions concerning the kernel function, the estimation was proved to be unbiased and consistent. The form of the kernel function in some cases was given by WATSON and LEADBETTER [24].

* Mathematical Research Institute, Hungarian Academy of Sciences, Budapest,

The other very generally used method of the nonparametrical density estimation is the so-called $k-NN$ method, that is, the k-th nearest neighbour method. Though the method originally was developed for pattern classification, it is not very difficult to determine its form for density function estimation. LOFTSGAARDEN and QUESENBERRY proved that the estimation was consistent [15]. There are some publications containing theorems proofs and concerning convergence and asymptotical unbiasedness [3, 6, 7, 11, 14, 18, 23, 25]. In spite of its popularity there are very few results concerning choice of suitable k. Some authors showed that in empirical methods, $k \approx \sqrt{N}$ the value gaves suitable results [2, 4, 15, 24]. FISCHER pointed out that in multidimensional cases the value of optimal k had to depend on the number of dimensions [6]. FUKUNAGA calculated analytically a form to determine the optimal k [9].

In the following part we shall introduce the main steps of this calculation, pointing out the consequences of some neglect, collecting the results of simulation.

We collected some published density function methods without describing them [1, 2, 8, 15, 16, 17, 19, 20, 21, 24].

1. The definition of k-NN estimation

Be (Ω, A, P) a probability space and $X = R^n$ the sample space. Let $X_1, X_2, \ldots, X_N$ be a set of N independent and identically distributed random vectors ($X_i \in X$, $\forall\ i = 1, \ldots, N$). Suppose that $f(x)$ exists as the density function of the sequence of $\{X_i\}_{i=1}^{\infty}$, it is continuous and twice differentiable.

The k-NN estimate of the density $f(X)$ at point X is formed as follows. Let d_k be the distance between X and its k-th nearest neighbour, distance being measured by any convenient metric $d(X, Y)$. Let $S(X)$ be the region about X containing its k-NN:

$$S(X) \equiv \{Y:\ d(X, Y) \leq d_k\} \tag{2}$$

and be $v(X)$ the volume of this region:

$$v(X) \equiv \int_{S(X)} dY\,. \tag{3}$$

Then the estimation of density $f(X)$ at point X_i:

$$\hat{f}(X_i) \equiv \frac{k-1}{N \cdot v(X_i)} \quad \forall\ i,\ \ i = 1, 2, \ldots, N\,. \tag{4}$$

LOFTSGAARDEN and QUESENBERRY showed that if $k = k(N)$ was chosen so that $k \to \infty$ and $k/N \to 0$ as $N \to \infty$, then (4) was an asymptotically un-

biased and consistent estimation of $f(X)$. The meaning of (4) in the one-dimensional case is illustrated in Fig. 1.

It is easy to show that for given N and n the estimation $f(X)$ is different for different values of k. Our main purpose is the optimization of $k = k(N,n,f(X))$ To this aim k is optimized with respect to the mean-square-error criterion:

$$J(X_i) = E\{[f(X_i) - \hat{f}(X_i)]^2\}; \quad \forall X_i \in X \tag{5}$$

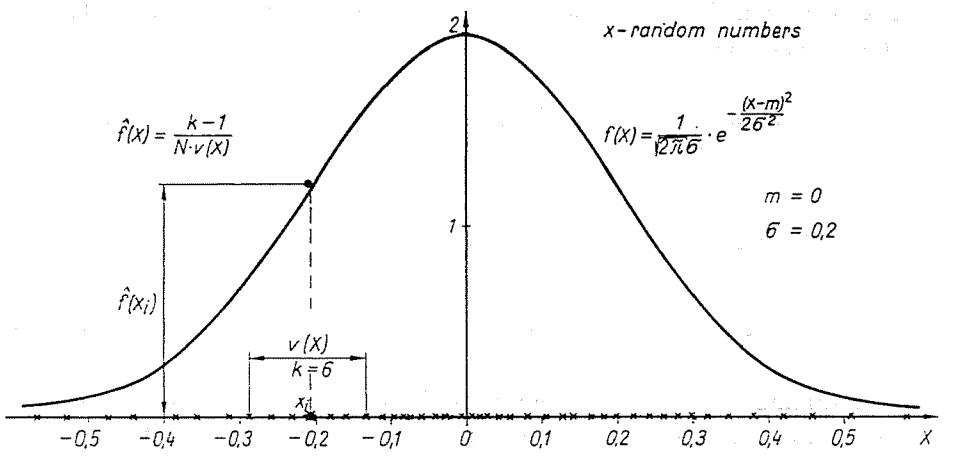

Fig. 1. Illustration of a density estimation

where the expectation is over the sample set $X_1, X_2, \ldots, X_N$. The integral mean-square-error

$$I(X) = \int_{\hat{X}} J(x)\, dX \tag{6}$$

Notice that also other criterion can be chosen to optimize k, for example:

$$I'(X) = \int_{\hat{X}} f(X) \cdot J(X)\, dX \tag{7}$$

Expressions (5) and (6) will be minimized with respect to (6) that is, the next necessary criterion:

$$\frac{\partial J(X)}{\partial k} = 0 \qquad \frac{\partial I(X)}{\partial k} = 0 \tag{8}$$

2. Volume to coverage relation

Using the k-NN estimator (6)

$$J(X) = \left(\frac{k-1}{N}\right)^2 \cdot E\left\{\frac{1}{v^2(X)}\right\} - 2 \cdot \frac{k-1}{N} \cdot E\left\{\frac{1}{v(X)}\right\} \cdot f(X) + f^2(X) \tag{9}$$

Thus, expressions for $E\{1/v^2(X)\}$ and $E\{1/v(X)\}$ are needed in terms of k. This is possible by expressing $v(X)$ in terms of the coverage of $S(X)$.

Be the probability of the domain of the k nearest neighbour (denoted $S(X)$) of the given point X_i:

$$u(X) \equiv \int_{S(X)} f(Y) \cdot dY = P\big(X \in S(X)\big) \tag{10}$$

It is known from literature that $u(X)$ has a beta distribution with the parameters $\beta(p, q)$, especially $p = k$ and $q = N - k + 1$ [10]:

$$f_u(X) = \frac{1}{B(p,q)} \cdot X^{p-1} \cdot (1 - X)^{q-1}; \quad p, q > 0, \quad 0 \leq X \leq 1 \tag{11}$$

in our case

$$f_u(X) = \frac{N!}{(k-1)! \cdot (N-k)!} \cdot u^{k-1} \cdot (1-u)^{N-k}, \quad 0 \leq u \leq 1. \tag{12}$$

The next step is to express $u(X)$ in terms of $v(X)$ in order to estimate the expectation of $u(X)$.

Assuming $f(X)$ to be continuous and its third derivate is to exist in the neighbourhood of X so it can be expanded in Taylor series about X:

$$u(X) \cong \int_{S(X)} \left\{ f(X) + \left[\frac{\partial f(X)}{\partial X}\right]^T \cdot (Y - X) + \frac{1}{2} \cdot (Y - X)^T \cdot \frac{\partial^2 f(X)}{\partial X^2} \cdot (Y - X) \right\} dY \tag{13}$$

and

$$\left[\frac{\partial f(X)}{\partial X}\right]^T \equiv \left[\frac{\partial f(X)}{\partial x_1}, \frac{\partial f(X)}{\partial x_2}, \ldots, \frac{\partial f(X)}{\partial x_n}\right] \tag{14}$$

$n \times 1$-vector

$$\frac{\partial^2 f(X)}{\partial X^2} \equiv \begin{bmatrix} \frac{\partial^2 f(X)}{\partial x_1^2}, & \frac{\partial^2 f(X)}{\partial x_1 \partial x_2}, & \ldots, & \frac{\partial^2 f(X)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(X)}{\partial x_2 \partial x_1}, & \frac{\partial^2 f(X)}{\partial x_2^2}, & \ldots, & \frac{\partial^2 f(X)}{\partial x_2 \partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f(X)}{\partial x_n \partial x_1}, & \frac{\partial^2 f(X)}{\partial x_n \partial x_2}, & \ldots, & \frac{\partial^2 f(X)}{\partial x_n^2} \end{bmatrix}$$

It seems that (13) is a matrix-vector equation but for simplicity their notions will not be distinguished only the superscript T indicates the matrix transpose operation. Although there is no condition for the metric, for the sake of sim-

plicity let us assume a conventional quadratic distance function such as:

$$d^2(X, Y) = (Y - X)^T \cdot A \cdot (Y - X) \tag{15}$$

Using the symmetry of the region $S(X)$, the second term in (13) is zero. Also using matrix relations $X^T \cdot A \cdot X = tr(A \cdot X \cdot X^T)$:

$$u(X) \cong f(X) \int_{S(X)} dY + \frac{1}{2} \cdot tr\left\{\left(\int_{S(X)} (Y - X) \cdot (Y - X)^T \cdot dY \right) \left(\frac{\partial^2 f(X)}{\partial X^2} \right)\right\}. \tag{16}$$

The first term of (16) is $f(X)$ $v(X)$. The integral of the second term can be calculated as the covariance matrix of a uniform distribution in an ellipsoidal region. The result can easily be calculated [17] (as the volume of n-dimensional hypersphere) as:

$$\int_{S(X)} (Y - X) \cdot (Y - X)^T dY = \frac{1}{(n+2)\pi} \cdot \Gamma\left(\frac{n+2}{2}\right) \cdot v^{1+\frac{2}{n}}(X) \cdot |A|^{1/n} \cdot A^{-1} \tag{17}$$

where $\Gamma(.)$ is the gamma function and

$$v(X) \equiv \int_{S(X)} dX = \frac{\pi^{n/2} \cdot d_k^n(X)}{|A|^{1/2} \cdot \Gamma\left(\frac{n+2}{2}\right)}. \tag{18}$$

Using the above equations (16) becomes

$$u(X) \approx f(X) \cdot v(X) + c(X) \cdot v^{1+2/n}(X) \tag{19}$$

where

$$c(X) = \frac{\Gamma^{2/n}\left(\frac{n+2}{2}\right)}{2 \cdot \pi \cdot (n+2)} \cdot tr\left\{ \left(\frac{A}{|A|^{1/n}} \right)^{-1} \cdot \frac{\partial^2 f(X)}{\partial X^2} \right\}. \tag{20}$$

We calculate a relation between $u(X)$ and $v(X)$ but we can not directly express $v(X)$ except the case $n = 2$. The first term of (16) is a good approximation to $u(X)$ and becomes even better as N increases and correspondingly $v(X) =$ gets smaller. Thus using approximation $u(X) \approx f(X) \cdot v(X)$

$$\frac{1}{v(X)} \approx \frac{f(X)}{u(X)} + \frac{c(X)}{f^{2/n}(X) \cdot u^{1-2/n}(X)} \tag{21}$$

We have to remark immediately that in this deduction this is the first neglect whose range of validity must be determined. (21) can be used as first approximation and its use simplifies the deduction.

3. Determination of the moments of u(X)

To establish the optimal k, expression, of $E\{1/v(X)\}$ and $E\{1/v^2(X)\}$ are needed. But because $1/v(X)$ is obtained in terms of $u(X)$ therefore the first step is to determine the moments of $u(X)$.

The distribution of $u(X)$ is known therefore the equation determining the moments:

$$E\{u(X)^\lambda\} = \frac{1}{B(k, N-k+1)} \cdot \int_0^1 u^{k+\lambda-1} \cdot (1-u)^{N-k} du = \frac{\Gamma(N+1) \cdot \Gamma(k+\lambda)}{\Gamma(N+1+\lambda) \cdot \Gamma(k)} \tag{22}$$

where $B(.,.)$ is the known beta function [5, 13, 22]:

$$B(x,y) = \frac{\Gamma(x) \cdot \Gamma(y)}{\Gamma(x+y)}, \quad B(k, N-k+1) = \frac{\Gamma(k) \cdot \Gamma(N-k+1)}{\Gamma(N+1)} \tag{23}$$

Using expression (22) for arbitrary values of the required moments are:

$$E\left\{\frac{1}{u(X)}\right\} = \frac{N}{k-1}, \qquad E\left\{\frac{1}{u^2(X)}\right\} = \frac{N \cdot (N-1)}{(k-1) \cdot (k-2)} \tag{24}$$

$$E\left\{\frac{1}{u^{1-2/n}(X)}\right\} = \frac{\Gamma(N+1) \cdot \Gamma\left(k-1+\frac{2}{n}\right)}{\Gamma\left(N+\frac{2}{n}\right) \cdot \Gamma(k)} \tag{25}$$

$$E\left\{\frac{1}{u^{2-4/n}(X)}\right\} = \frac{\Gamma(n+1) \cdot \Gamma\left(k-2+\frac{4}{n}\right)}{\Gamma\left(N-1+\frac{4}{n}\right) \cdot \Gamma(k)} \tag{26}$$

Although the exact values are given by these expressions in the above mentioned form if seems to be hopeless to use the expressions of the expectations $E\{1/v(X)\}$ and $E\{1/v^2(X)\}$ for determining the optimal value of k. The greatest difficulty is due to the determination of terms $2/n$ and $4/n$ acting in the argument of the factorial function. While these are no integer values in the case of $n > 3$ and therefore expressions (25) and (26) cannot be simplified the value of $k(N)$ analytically expressed cannot be after the derivation needed to calculate the optimum because these terms stay in the argument of the factorial function.

To get rid of the above difficulties, FUKUNAGA [9] chose the following approximation instead of the exact expression of the moment (22):

$$E\left\{\left(\frac{1}{u(X)}\right)^\lambda\right\} = \left(\frac{k-1}{N}\right)^\lambda \tag{27}$$

It must be mentioned that this approximation gives a good result only asymptotically for $n \to \infty$.

Therefore it can be predicted that the formula in course of deduction will be right only at higher dimensions. This will be the case. For $N \gg k$ the following approximation [22] is valid:

$$\frac{\Gamma(x)}{\Gamma(x+a)} \approx x^{-a} \tag{28}$$

$$E\{u(X)^{\lambda}\} \approx (N+1)^{\lambda} \cdot \frac{\Gamma(k+\lambda)}{\Gamma(k)} \tag{29}$$

Unfortunately this is an analytical result from approximation (27).

4. The optimization of k

Our criterion in determining the optimal k is to minimize the functional (6). Expressions (27) and (21) quantify (9):

$$E\left\{\frac{1}{v(X)}\right\} = f(X) \cdot E\left\{\frac{1}{u(X)}\right\} + E\left\{\frac{1}{u^{1-2/n}(X)}\right\} \cdot c(X) \cdot f^{-2/n}(X) \tag{30}$$

$$\begin{aligned} E\left\{\frac{1}{v^2(X)}\right\} = f^2(X) \cdot E\left\{\frac{1}{u^2(X)}\right\} + 2 \cdot E\left\{\frac{1}{u^{2-2/n}(X)}\right\} \cdot c(X) \cdot f^{1-2/n}(X) + \\ + E\left\{\frac{1}{u^{2-4/n}(X)}\right\} \cdot c^2(X) \cdot f^{-4/n}(X) . \end{aligned} \tag{31}$$

Substituting in (9) we obtain

$$J(X) = \frac{1}{k} \cdot f(X) + c^2(X) f^{-4/n}(X) \cdot \left(\frac{k}{n}\right)^{4/n} . \tag{32}$$

This expression shows that the value of k can indeed be ∞ and k/N may tend to zero as it is required by asymptotical unbiasedness and consistency of estimation (nearest neighbour). Differentiating with respect to k, to minimize the expression, we obtain

$$k_0(N) = \left[\frac{n \cdot f^{2+4/n}(X)}{4 \cdot c^2(X)}\right]^{\frac{n}{n+4}} \cdot N^{\frac{4}{n+4}}, \tag{33}$$

It should be noted that this is not the final result as only the minimum of functional (5) referring to fixed X has been determined. Anyhow it can be

stated that $k_0(N) \to \infty$, $k_0(N)/N \to 0$ and $J_0(X) \to 0$ when the dimension increases. The optimal $k_0(N)$ seems to depend on the dimension of feature space and on the basic distribution. $c(X)$ defined in (20) can also be interpreted as the measure of the unformity of the underlying distribution. Thus, if $c(X)$ is large (i.e. $f(X)$ has a large second derivate), then it changes very rapidly in the region about X and therefore a lower value of k should be chosen. This is shown by formula (33) as well.

In order to minimize the functional (6), i.e. the mean-square-error of estimation of total function $f(X)$ to be minimum, because of having $I(X)$ instead of $J(X)$, the appropriate terms must be integrated in expression (33):

$$k_0(N) = \left[\frac{n \cdot \int f^{2+4/n}(X)dX}{4 \cdot \int c^2(X)dX}\right]^{\frac{n}{n+4}} \cdot N^{\frac{4}{n+4}} \tag{34}$$

More exactly:

$$k_0(N) = \left[\frac{n \cdot (n+2)^2 \cdot \pi^2 \cdot \int f^2(X)dX}{\Gamma^{4/n}\left(\frac{n+2}{2}\right) \cdot \int f^{-4/n}(X) \cdot tr^2\left\{\left[\frac{A}{|A|^{1/n}}\right]^{-1} \frac{\partial^2 f(X)}{\partial X^2}\right\} dX}\right]^{\frac{n}{n+4}} \cdot N^{\frac{4}{n+4}}. \tag{35}$$

This expression yields the expression of optimal k with the given neglects.

FUKUNAGA [9] has shown that expression (35) is invariant using linear transformation. This is important, because the values of optimal k can be generalized to a certain extent in the case of a given distribution. If a density function f(X) is given, which has expected value M and covariance matrix K, then on the basis of the system of eigenvalues — eigenvectors of K, it is always possible to give a transformation, making K a diagonal or identity matrix. While during the transformation the functionals $J(X)$ and $I(X)$ are multiplied by a constant, this is not enough for their minimum to change, i.e. the expression of optimal $k_0(N)$ remains unchanged both in the original space and in the transformed space:

$$k_0^{(X)}(N) = k_0^{(Z)}(N) \tag{36}$$

where subscripts refer to individual spaces.

5. The case of Gaussian distribution

As an example let us consider random variables $\{X_i\}_{i=1}^{\infty}$, deriving from population of normal, Gaussian distribution, having expected value M_x and covariance matrix K_x. Performing the linear transformation, which makes K diagonal, in the new space Z, A_z will equal I (identity matrix) and K_z will

equal I, where A_z comes from (15) and K_z is the new covariance matrix. Taking all these into consideration, replacing them in (35) and performing the integrations, we obtain:

$$k_0(N) = \left[\frac{n \cdot (n+2)^2 \cdot \Gamma^{-4/n} \left(\frac{n+2}{2} \right) \cdot \left(\frac{n-2}{2} \right)^{2+\frac{n}{2}}}{n^2 - 6n + 16} \right]^{\frac{n}{n+4}} \cdot N^{\frac{4}{n+4}}. \qquad (37)$$

Seemingly, this expression is made up of two well separable parts. Therefore it can be written in the following form:

$$k_0(N) = f(n) \cdot N^{\frac{4}{n+4}} \qquad (38)$$

The first factor of the expression depends on the dimension of feature space, only the second factor depends on the sample number as well. As a matter of curiosity, the function $f(n)$ has been tabulated by a computer. The result is shown in Fig. 2.

It can be ascertained on the basis of Fig. 2. and expression (37) that the result is not valid in the case of dimensions 1 and 2, further there are large neglects up to dimension 7. too. The second factor of expression (38) increases monotonically, when the dimension increases, therefore function $f(n)$ must also increase monotonically, so that $k_0(N)$ has to be increased monotonically. But on the basis of Fig. 2, $f(n)$ seems to have a maximum, hence, there exists an optimal dimension (on the basis of the results this would be 6), but this is physically impossible. *Without making any postulate on the dimension at the beginning of the demonstration, it is absolutely unlinely that certain physical*

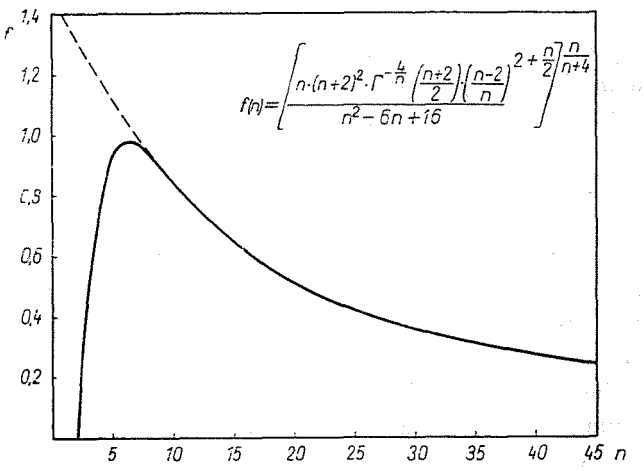

Fig. 2. Function $f(n)$

Table 1

The optimal value of $k_0(N)$ in case of multidimensional Gaussian distribution calculated by expression (37)

N sample size	$n = 3$	$n = 5$	$n = 7$	$n = 10$	$n = 12$
	number of dimensions				
100	6.28	7.10	5.17	3.16	2.40
200	9.33	9.66	6.66	3.86	2.86
300	11.77	11.57	7.72	4.33	3.16
400	13.87	13.15	8.57	4.70	3.40
500	15.75	14.52	9.29	5.01	3.59
600	17.48	15.74	9.93	5.28	3.76
700	19.09	16.86	10.50	5.52	3.90
800	20.61	17.89	11.03	5.73	4.04
900	22.04	18.85	11.51	5.93	4.16
1000	23.41	19.75	11.96	6.11	4.27
1100	24.72	20.61	12.38	6.27	4.38
1200	25.98	21.42	12.78	6.43	4.47
1300	27.19	22.20	13.16	6.58	4.56
1400	28.37	22.94	13.52	6.72	4.65
1500	29.51	23.66	13.86	6.86	4.73
1600	30.62	24.34	14.19	6.98	4.80
1700	31.70	25.00	14.50	7.10	4.88
1800	32.75	25.65	14.80	7.22	4.95
1900	33.78	26.27	15.10	7.34	5.02
2000	34.78	26.88	15.39	7.44	5.08

processes take place more optimally in a space with, let us say, dimonsion 6, than in one with for example dimension 3. This also shows that the mathematical neglects strongly restrict the validity of the result. On Fig. 2 the expectable form of the curve $f(n)$ is traced with dotted line in case of lower dimensions. It should be noted, however, that since in case of $n = 2$ the reciprocals of the moments should be replaced into expressions (25) and (26) (i.e. instead of $1/u(X)$ we put $u(X)$ with their appropriate power), therefore it seems very likely that during the mathematical demonstration it is suitable to separate the cases of low and high dimensions.

We also note in the mentioned article of FUKUNAGA [9] there is a wrong value among the presented values of the function $f(n)$ as well.

We calculated the optimal values of k_0 for some cases with the help of the obtained result and we presented them in tabulated form. Ignoring the neglects, it is very likely that if the dimension is high and the sample size is

the same, then a smaller value of k_0 should be chosen. This is seen in the presented table as well.

However, we can make a few remarks of general validity, according to the above table. First of all, k_0 results in decimal numbers too, but there integral parts have to be taken into account. The values of k_0 in the table are accurate in order to point out the dependence on the number of the samples. Increasing the number of samples, the value of k_0 is seen to change very slowly. On the other part, the change will decrease by increasing the number of dimension, the measure of this change is so large that it has to be taken into account. It cannot be settled if it is a right conclusion or consequence of neglects. Using computer simulation further results can be obtained.

Further interesting conclusion is if $n = 3$ (the number of dimensions), the optimal value of k_0 is less then $k_0 = \sqrt[]{N}$ suggested in the literature, but the difference is small. In connection with Fig. 2 it has been remarked if the number of dimension is less then a higher value can be expected, this observation supports the use of approximation $\sqrt[]{N}$ or a higher value.

6. Results of simulations

It is practical to make computer simulations for controlling the theoretical results. An one dimensional, standard Gaussian random number series were generated by computer then the unknown density function was estimated by the k-nearest neighbour method. The underlying distribution was known so we could calculate the mean-square-error of the estimation. For the calculation a computer program was written in ALGOL-60 language, run on a computer type RAZDAN-3 of the University Computing Centre.

The expected value and variance of the underlying distribution was zero and unit, respectively. These obligations do not change the generality of the results. We made several runs with distinct values of k and N (sample size). The results are shown in Fig. 3, the mean-square-error versus k (number of neighbours), the sample size is parameter. *It is very important and interesting that the curves have minima — as we have expected — versus k. Increasing N the sharpness of the minima decreases but exists! It is easy to calculate that the optimal value of k concerning the minima is higher than the value to be calculated from the approximation $\sqrt[]{N}$!*

In Fig. 4, the absolute error of the estimation is shown. The trend of the curves is similar to that mentioned above.

We have to note that if the variance of the underlying distribution is not unity the mean-square-error will be higher or lower (either $\sigma < 1$ or $\sigma > 1$), but the minimum is at the same region. This is very interesting because the simulation results can be used in the case of optional $N(m, \sigma)$ distribution after a linear transformation, as mentioned before.

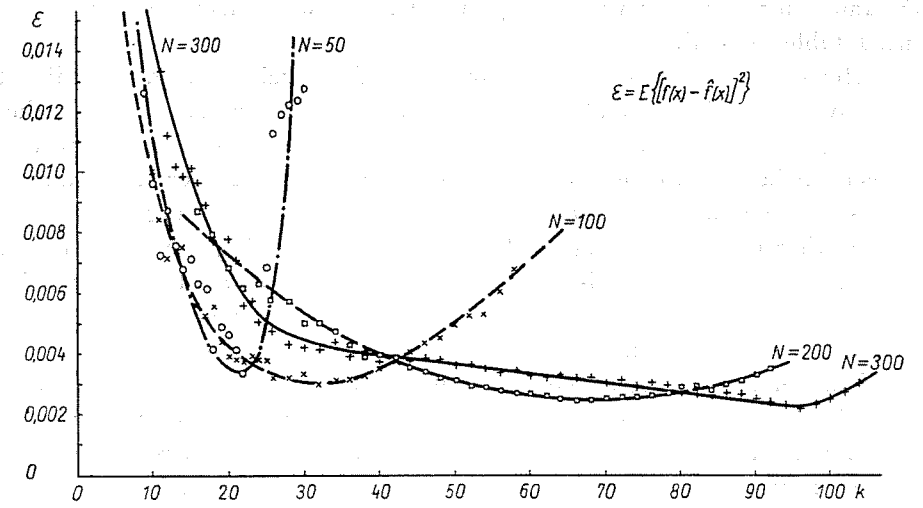

Fig. 3. The mean-square-error of the estimation, the underlying distribution was standard Gaussian with parameters m = 0 and σ=1. Parameter the sample size N

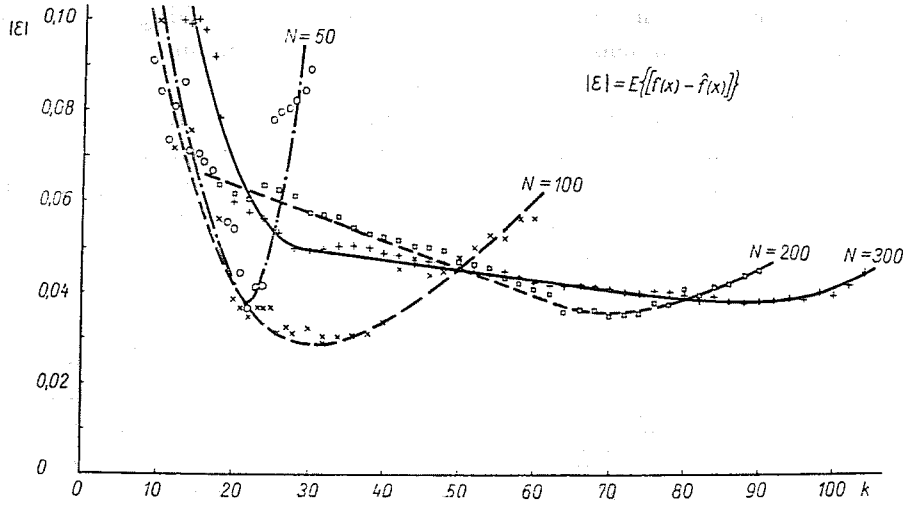

Fig. 4. The absolute error of the estimation, the underlying distribution was standard Gaussian with parameters $m = 0$ and $\sigma = 1$. Parameter is the sample size N

Summary

After a short survey the density estimation, declaration of the k-nearest neighbour method is given. The aim of this paper is to give a possible optimum choice of k according to FUKUNAGA and to point out the consequences of its neglect and to collect the results. After the theoretical treatment, simulation results are presented.

References

1. ANDREWS, H. C.: Introduction to Mathematical Techniques in Pattern Recogintion. John Wiley, New York, (1972).
2. CAUCOLLOS, T.: Estimation of a Multivariate Density. Ann. Inst. Stat. Math., **18**, pp. 179—189, (1966).
3. COVER, T. M.—HART, P. E.: Nearest Neighbor Pattern Classification. IEEE Trans. Theory, **IT-13**, pp. 21—27, (1967).
4. EPANECHNIKOV, V. A.: Nonparametric Estimation of a Multivariate Probability Density. Theory Prob. Appl. (USSR), **14**, pp. 153—158, (1969).
5. FAZEKAS, F.—FREY, T.: Operátorszámítás. Speciális függvények. Műszaki Matematikai Gyakorlatok, C. I. főszerk.: Fazekas, F. Tankönyvkiadó, Budapest, (1965).
6. FISCHER, F. P.: k-Nearest Neighbor Rules. Ph. D. dissertation, School Elec. Eng., Purdue University, Lafayette, Indiana, (1971).
7. FRITZ, J.: Distribution Free Exponential Error Bound for Nearest Neighbor Pattern Classification. (submitted to IEEE trans. Inf. Theory)
8. FUKUNAGA, K.: Introduction to Statistical Pattern Recognition. Academic Press, New York, (1972).
9. FUKUNAGA, K.—HOSTETLER, L. D.: Optimization of k-Nearest Neighbor Density Estimates. IEEE Trans. Inf. Th., **IT-19**, pp. 320—326, (1973).
10. GNEDENKO, B. V.: The Theory of Probability MIR Publishers, Moscow, 2nd Ed., (1973).
11. GÁBOR, GY.: A "nearest neighbor" osztályozási módszer. TKI Szem. Közl., (1971).
12. GJUNTER, N. M.—KUZMIN, R. O.: Felsőbb matematikai példatár. Tankönyvkiadó, Budapest, (1952).
13. JAHNKE—EMDE, F.: Tables of Highers Functions. Teubner Verlagsgesellschaft, Leipzig, (1952).
14. LEVINE, A.—LUSTICK, L.—SALTZBERG, B.: The Nearest-Neighbor Rule for Small Samples Drawn from Uniform Distribution. IEEE Trans. Inf. Th., **IT-19**, pp. 697—699, (1973).
15. LOFTSGAARDEN, D. O.—QUESENBERRY, C. P.: A Nonparametric Estimate of a Multivariate Density Function. Ann. Math. Statist., **36**, pp. 1049—1051, (1965).
16. MEISEL, W. S.: Computer-Oriented Approaches to Pattern Recognition. Academic Press, New York, (1972).
17. MILLER, K. S.: Multidimensional Gaussian Distribution. John Wiley, New York, (1964).
18. PETERSON, D. W.: Some Convergence Properties of a Nearest Neighbor Decision Rule. IEEE Trans. Inf. Th., **IT-16**, pp. 26—31, (1970).
19. REITSMA, A.: On Approximations to Sampling Distributions of the Mean for Samples from Non-Normal Populations. Ann. Math. Statist., **34**, pp. 1308—1314, (1962).
20. REJTŐ, L.—RÉVÉSZ, P.: Density Estimation and Pattern Recognition. Problems of Control and Information Theory, **2**, pp. 67—80, (1973).
21. ROSENBLATT, M.: Remarks on Some Nonparametric Estimates of a Density Function. Ann. Math. Statist., **27**, pp. 832—837, (1956).
22. TITCHMARSCH, E. C.: The Theory of Functions. Oxford University Press, Oxford, (1939).
23. WAGNER, T. J.: Convergence of the Nearest Neighbor Rule. IEEE Trans. Inf. Th., **IT-17**. pp. 566—571, (1971).
24. WATSON, G. S.—LEADBETTER, M. R.: On the Estimation of the Probability Density. Ann. Math. Statist., **34**, pp. 480—491, (1962).
25. WOLVERTON, Ch.: Strong Consistency of an Estimate of the Asymptotic Error Probability of the Nearest Neighbor Rule.
IEEE Trans. Inf. Th., **IT-19**, pp. 119—120, (1973).

Dr. Miklós VAJTA Jr.
Dr. József FRITZ
} H-1521 Budapest