Thus, the limiting distribution of the square error of a parametric estimate of a multidimensional normal density is given by the relation

$$\lim_{n\to\infty} \mathbf{P}\{n2^{k+3}\pi^{k/2}\sqrt{\det C}\Phi_n < x\} = F(x).$$

Received by the editors
June 10, 1968

REFERENCES

[1] T. ANDERSON, *Introduction to Multi-dimensional Statistical Analysis*, Fizmatgiz, Moscow, 1963. (In Russian.)
[2] G. M. MANIYA, *Quadratic error in estimating multidimensional normal distribution densities from sample data*, Theory Prob. Applications, 13 (1968), pp. 341–343.
[3] J. GURLAND, *Quadratic forms in normally distributed random variables*, Sankhyā, 17 (1956), pp. 37–50.

NON-PARAMETRIC ESTIMATION OF A MULTIVARIATE PROBABILITY DENSITY

V. A. EPANECHNIKOV

(*Translated by B. Seckler*)

Introduction

Let

$$X_i = X(x_1^{(i)}, x_2^{(i)}, \cdots, x_k^{(i)}), \qquad i = 1, \cdots, n,$$

be a given sample of n independent realizations of a k-dimensional random variable $X(x_1, x_2, \cdots, x_k)$ from a population characterized by a continuous k-variate probability density $f(x_1, \cdots, x_k)$. We define the multivariate empirical probability density $f_n(x_1, \cdots, x_k)$ to be the function of sample values X_i given by

$$f_n(x_1, \cdots, x_k) = \frac{1}{n}\sum_{i=1}^{n}\prod_{l=1}^{k}\frac{1}{h_l(n)}K_l\left(\frac{x_l - x_l^{(i)}}{h_l(n)}\right). \tag{1}$$

Each "kernel" $K_l(y)$ has the following properties:

(2)
(a) $0 \leqq K_l(y) < C < \infty$,

(b) $K_l(y) = K_l(-y)$,

(c) $\int_{-\infty}^{\infty} K_l(y)\,dy = 1$,

(d) $\int_{-\infty}^{\infty} K_l(y)y^2\,dy = 1$,

(e) $\int_{-\infty}^{\infty} K_l(y)y^m\,dy < \infty$ for $0 \leqq m < \infty$,

and the "spreading" coefficients $h_l(n)$ of the kernels depend in general on the sample size n and tend to zero as $n \to \infty$.

Non-parametric estimation of a true univariate ($k = 1$) probability density of the form (1) was considered by Parzen [1] with arbitrary kernel $K(y)$ and by Rosenblatt [2] (for $k = 1$) and Maniya [3] (for $k = 2$) with a specific kernel of the form

$$K(y) = \begin{cases} a & \text{for } |y| \leqq 1/2a, \\ 0 & \text{for } |y| > 1/2a. \end{cases}$$

A bivariate empirical probability density of the form

$$f_n(x_1, x_2) = \frac{1}{nh^2(n)} \sum_{i=1}^{n} H\left(\frac{x_1 - x_1^{(i)}}{h(n)}; \frac{x_2 - x_2^{(i)}}{h(n)}\right)$$

was used by Nadaraya [4].

This paper examines some properties of the multivariate empirical probability density (1) with kernels of arbitrary form (subject to restrictions (2)) in the case where the true probability density has a Taylor expansion in all its arguments about each point $X(x_1, \cdots, x_k)$.

1. Asymptotic Properties of the Empirical Probability Density

Performing some simple operations in the expression for the averaged empirical density,

$$\mathbf{E}f_n(x_1, \cdots, x_k) = \int \cdots \int \left[\prod_{l=1}^{k} \frac{1}{h_l(n)} K\left(\frac{x_l - y_l}{h_l(n)}\right)\right] f(y_1, \cdots, y_k)\, dy_1 \cdots dy_k,$$

we can write

$$\mathbf{E}f_n(x_1, \cdots, x_k) = \int \cdots \int \left[\prod_{l=1}^{k} K_l(y_l)\right] f(x_1 + h_1 y_1, \cdots, x_k + h_k y_k)\, dy_1 \cdots dy_k. \tag{3}$$

We expand $f(x_1 + h_1 y_1, \cdots, x_k + h_k y_k)$ in a Taylor series with respect to all x_i about the point $X(x_1, x_2, \cdots, x_k)$, we integrate the right-hand side of (3) and we let $n \to \infty$. This yields for the averaged difference between the empirical and true probability densities, $\mathbf{E}\Delta f_n(x_1, \cdots, x_k) = \mathbf{E}[f_n(x_1, \cdots, x_k) - f(x_1, \cdots, x_k)]$, the result

$$\mathbf{E}\Delta f_n(x_1, \cdots, x_k) \sim \frac{1}{2} \sum_{l=1}^{k} \frac{\partial^2 f(x_1, x_2, \cdots, x_k)}{\partial x_l^2} h_l^2(n). \tag{4}$$

The mean-square error of approximation defined by

$$\mathbf{E}(\Delta f_n(x_1, \cdots, x_k))^2 = \mathbf{E}[f_n(x_1, \cdots, x_k) - f(x_1, \cdots, x_k)]^2,$$

is equal to

$$\begin{aligned}
\mathbf{E}(\Delta f_n(x_1, \cdots, x_k))^2 = {} & \frac{1}{n^2}\left[\prod_{l=1}^{k} \frac{1}{h_l^2(n)}\right] \cdot \left\{ n\left[\prod_{l=1}^{k} h_l(n)\right] \right. \\
& \times \int \cdots \int \left[\prod_{l=1}^{k} K_l^2(y_l)\right] f(x_1 + h_1 y_1, \cdots, x_k + h_k y_k)\, dy_1 \cdots dy_k + n(n-1) \\
& \left. \times \left[\prod_{l=1}^{k} h_l^2(n)\right]\left[\int \cdots \int \left[\prod_{l=1}^{k} K_l(y_l)\right] f(x_1 + h_1 y_1, \cdots, x_k + h_k y_k) dy_1 \cdots dy_k\right]^2 \right\} \\
& - 2f(x_1, \cdots, x_k) \int \cdots \int \left[\prod_{l=1}^{k} K_l(y_l)\right] \\
& \times f(x_1 + h_1 y_1, \cdots, x_k + h_k y_k)\, dy_1 \cdots dy_k + f^2(x_1, \cdots, x_k).
\end{aligned} \tag{5}$$

Substituting the Taylor expansion for $f(x_1 + hy_1, \cdots, x_k + hy_k)$ about the point $X(x_1, \cdots, x_k)$ into the right-hand side of (5) and letting $n \to \infty$, we obtain

$$\begin{aligned}
\mathbf{E}(\Delta f_n(x_1, \cdots, x_k))^2 \sim {} & \frac{1}{n} f(x_1, \cdots, x_k) \prod_{l=1}^{k} \left[\frac{1}{h_l(n)} \int_{-\infty}^{\infty} K_l^2(y)\, dy\right] \\
& + \frac{1}{4}\left[\sum_{l=1}^{k} \frac{\partial^2 f(x_1, \cdots, x_k)}{\partial x_l^2} h_l^2(n)\right]^2.
\end{aligned} \tag{6}$$

Expressions (4) and (5) imply that when $h_l(n) \to 0$ and $n \prod_{l=1}^{k} h_l(n) \to \infty$ the empirical probability density (1) is a consistent estimator of the true probability density $f(x_1, \cdots, x_k)$ at each point $X(x_1, \cdots, x_k)$.

2. Relative Global Approximation Error

The relative global approximation error $\hat{u}^2$ due to approximating the true probability density $f(x_1, \cdots, x_k)$ by the empirical density $f_n(x_1, \cdots, x_k)$ is defined by[1]

$$\hat{u}^2 = \frac{1}{Q}\int \cdots \int \mathbf{E}(\Delta f_n(x_1, \cdots, x_k))^2 \, dx_1 \cdots dx_k, \tag{7}$$

where

$$Q = \int \cdots \int f^2(x_1, \cdots, x_k) \, dx_1 \cdots dx_k. \tag{8}$$

We arrive at an asymptotic relative global error (as $n \to \infty$) by substituting the value of $\mathbf{E}(\Delta f_n(x_1, \cdots, x_k))^2$ from (6) into the right-hand side of (7):

$$\hat{u}^2 \sim \frac{1}{Q}\left[\frac{1}{n}\prod_{l=1}^{k}\frac{1}{h_l(n)}\int_{-\infty}^{\infty} K_l^2(y)\, dy + \frac{1}{4}\int \cdots \int \left[\sum_{l=1}^{k} \frac{\partial^2 f(x_1, \cdots, x_k)}{\partial x_l^2} h_l^2(n)\right]^2 dx_1 \cdots dx_k\right].$$

Setting $K_l(y) = K(y)$ and $h_l(n) = h(n)$, we shall examine ways of minimizing the relative global error

$$\hat{u}^2 \sim \frac{n^{-1}h^{-k}(n)L^k + (\frac{1}{4})h^4(n)M}{Q}, \tag{9}$$

where

$$L = \int_{-\infty}^{\infty} K^2(y)\, dy, \tag{10}$$

$$M = \int \cdots \int \left[\sum_{l=1}^{k} \frac{\partial^2 f(x_1, \cdots, x_k)}{\partial x_l^2}\right]^2 dx_1 \cdots dx_k. \tag{11}$$

a) Optimization of the spreading coefficient. To determine the optimum spreading coefficient $h(n) = h_0(n)$ minimizing the asymptotic relative global error $\hat{u}^2$ (as $n \to \infty$), we differentiate the right-hand side of (9) with respect to $h(n)$ and we equate the derivative to zero. Thus for $n \to \infty$,

$$h_0(n) \sim \left(\frac{kL^k}{nM}\right)^{1/(k+4)}. \tag{12}$$

b) Optimization of the form of the kernel. From (9) it follows that to determine the optimum kernel form $K(y) = K_0(y)$ minimizing the relative global error, it suffices to minimize the expression $L = \int K^2(y)\, dy$ for fixed h, n and k subject to the additional conditions (2b)–(2d). This problem belongs to those isoperimetric problems of the calculus of variations with constraints present (see, for example, [5]).

Euler's equation for such variational problems can be written in the form $K(y) + \lambda_1 + \lambda_2 y^2 = 0$, where the parameters λ_1 and λ_2 are determined from conditions (2b)–(2d). Having

[1] This error could also be termed the relative total mean-square error. The relative global error $\hat{u}^2$ is a special case of the relative weighted total mean-square error, with weight function $\psi(x_1, \cdots, x_k)$, defined by

$$\frac{1}{Q}\int \cdots \int \mathbf{E}(\Delta f_n(x_1, \cdots, x_k))^2 \psi(x_1, \cdots, x_k)\, dx_1 \cdots dx_k.$$

ascertained constants λ_1 and λ_2, we find the optimum kernel $K_0(y)$ to be

$$K_0(y) = \begin{cases} \dfrac{3}{4\sqrt{5}} - \dfrac{3y^2}{20\sqrt{5}} & \text{for } |y| \leqq \sqrt{5}, \\ 0 & \text{for } |y| > \sqrt{5}. \end{cases} \tag{13}$$

The resultant optimum kernel $K_0(y)$ is independent of the true probability density, the sample size and the dimensionality of the space.

TABLE 1

$K(y)$	L	r
$K_0(y)$ (13)	$3/5\sqrt{5}$	1
$\frac{\sqrt{\pi^2-8}}{4}\cos\frac{\sqrt{\pi^2-8}}{2}y$ for $\lvert y\rvert \leqq \frac{\pi}{\sqrt{\pi^2-8}}$; 0 for $\lvert y\rvert > \frac{\pi}{\sqrt{\pi^2-8}}$	$\pi\sqrt{\pi^2-8}/16$	1.001
$1/\sqrt{6} - \lvert y\rvert/6$ for $\lvert y\rvert \leqq \sqrt{6}$; 0 for $\lvert y\rvert > \sqrt{6}$	$\sqrt{6}/9$	1.015
$(2\pi)^{-1/2}e^{-y^2/2}$	$1/2\sqrt{5}$	1.051
$1/2\sqrt{3}$ for $\lvert y\rvert \leqq \sqrt{3}$; 0 for $\lvert y\rvert > \sqrt{3}$	$1/2\sqrt{3}$	1.077
$\frac{1}{2}e^{-\sqrt{2}\lvert y\rvert}$	$1/4\sqrt{2}$	1.320

Table 1 gives the values of the integral $L = \int_{-\infty}^{\infty} K^2(y)\,dy$, occurring in (9), and of the ratio $r = \int_{-\infty}^{\infty} K^2(y)\,dy \Big/ \int_{-\infty}^{\infty} K_0^2(y)\,dy$ for certain kernels $K(y)$.

c) COMBINED OPTIMIZATION OF THE KERNEL AND SPREADING COEFFICIENT. To minimize the relative global error $\hat{u}^2$ completely, one has to minimize $\hat{u}^2$ first with respect to the form of the kernel $K(y)$ and then with respect to the spreading coefficient $h(n)$. Substituting the optimum $h_0(n)$ from (12) and $K_0(y)$ from (13) into the right-hand side of (9), we obtain, as $n \to \infty$,

$$\hat{u}^2_{\min\min} = \hat{u}_0^2 \sim \frac{(k+4)(3/5\sqrt{5})^{4k/(k+4)}M}{4n^{4/(k+4)}k^{k/(k+4)}Q}, \tag{14}$$

where Q and M are given by (8) and (11).

Table 2 cites the values of the integral M occurring in the expressions (9) and (14) for $\hat{u}^2$ for certain true probability densities $f(x_1, \cdots, x_k)$.

TABLE 2

$f(x_1, \cdots, x_k)$	M
$\dfrac{1}{(2\pi)^{k/2}} \exp\left(-\dfrac{1}{2}\sum_{p=1}^{k} x_p^2\right)$	$\dfrac{k(2+k)}{4(2\sqrt{\pi})^k}$
$\left(\dfrac{2}{\pi}\right)^k \prod_{p=1}^{k} \dfrac{1}{(x_p+1)^2}$	$\dfrac{k(98k+307)}{4(2\sqrt{\pi})^k}$
$\dfrac{1}{\pi^k} \prod_{p=1}^{k} \dfrac{1}{x_p^2+1}$	$\dfrac{k(5+k)}{4(2\pi)^k}$

d) VARIANCE OF THE RELATIVE TOTAL SQUARE ERROR. To estimate the random deviation of u^2_{samp}, the sampling value of the relative total square error, from the relative global error $\hat{u}^2$, we consider the asymptotic variance ($n \to \infty$) of the relative total square error

$$\begin{aligned}\sigma^2_{u^2} = \mathbf{E}(u^2)^2 - (\hat{u}^2)^2 = Q^{-2}\Big\{\mathbf{E}\Big[\int \cdots \int [f_n(x_1, \cdots, x_k) - f(x_1, \cdots, x_k)]^2\, dx_1 \cdots dx_k\Big]^2 \\ - \Big[\int \cdots \int \mathbf{E}[f_n(x_1, \cdots, x_k) - f(x_1, \cdots, x_k)]^2\, dx_1 \cdots dx_k\Big]^2\Big\}.\end{aligned} \tag{15}$$

Integrating on the right-hand side of (15) and discarding higher order terms (the tedious intermediary calculations have been omitted), we obtain as $n \to \infty$,

$$\sigma^2_{u^2} \sim 2\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} K(y)K(y-z)\,dy\right]^2 dz/nh^k(n)Q. \tag{16}$$

For the optimum values $h(n) = h_0(n)$ and $K(y) = K_0(y)$, we have

$$\sigma^2_{u_0^2} \sim \frac{2(0.0875\sqrt{5})^k M^{k/(k+4)}}{n^{(8+k)/(4+k)} k^{k/(k+4)} Q} \tag{17}$$

(see (8), (11)).

Expressions (16) and (17) imply that the asymptotic ratio ($n \to \infty$) of the standard deviation σ_{u^2} of the relative total square error to the relative global error $\hat{u}^2$ is proportional to $\sigma_{\hat{u}^2}/\hat{u}^2 \sim n^{-k/2(k+4)}$ and tends to zero.

3. Determination of the Sample Size Assuring a Prescribed Level for the Minimum Relative Global Error

Solving equation (7) for n, we find how the required sample size depends on the admissible value of the minimum relative global error $\hat{u}^2_{\min}$.

Table 3 gives values of the same size that assure a prescribed level for the minimum relative global error $\hat{u}^2_{\min}$ when the true density is $f(x_1, \cdots, x_k) = (2\pi)^{-k/2}\exp(-\frac{1}{2}\sum_{l=1}^{k} x_l^2)$ and the kernel is taken to be of the form $K(y) = (2\pi)^{-1/2} e^{-(1/2)y^2}$.

TABLE 3

k \ $\hat{u}^2_{\min}$	0.1	0.2	0.3	0.4	0.5
1	22	11	6	4	3
2	58	21	11	7	5
3	175	52	26	16	11
4	600	150	67	38	24
5	$2.22 \cdot 10^3$	470	190	98	59

The values of the sample sizes cited in Table 3 (which were rounded off to the nearest higher integer) were obtained from the exact expression for the minimum relative global error

$$\hat{u}^2_{\min} = \frac{1}{n}\left\{\frac{1}{h_0^k(n)} + \frac{n-1}{[1+h_0^2(n)]^{k/2}} - \frac{n \cdot 2^{(k+2)/2}}{[2+h_0^2(n)]^{k/2}} + n\right\},$$

the optimum spreading coefficient $h_0(n)$ being related to the sample size n by

$$n = \frac{[(1+h_0^2(n))/h_0^2]^{(k+2)/2} - 1}{[(1+h_0^2(n))/(1+\frac{1}{2}h_0^2(n))]^{(k+2)/2} - 1}.$$

For small $\hat{u}^2$, we obtain

$$h_0(n) \sim \left[\frac{4}{n(k+2)}\right]^{1/(k+4)}, \qquad n \sim \frac{(k+4)^{(k+4)/4} \cdot (k+2)^{k/4}}{(\hat{u}^2)^{(k+4)/4} \cdot 2^{k+2}}.$$

The above non-parametric estimation of a true multivariate probability density can also be applied to solve various problems involving the statistical tests of hypotheses.

Received by the editors
February 1, 1967

REFERENCES

[1] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist., 35 (1962), pp. 1065–1076.
[2] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*, Ann. Math. Statist., 27 (1956), pp. 642–669.
[3] G. M. Maniya, *Remarks on nonparametric estimates of a bivariate probability density*, Soobshch. Akad. Nauk Gruz. SSR, 27, 4 (1961), pp. 385–400.
[4] E. A. Nadaraya, *Estimation of a bivariate probability density*, Soobshch. Akad. Nauk Gruz. SSR, 36, 2 (1964), pp. 267–268.
[5] R. Bellman, I. Glicksberg and O. Gross, *Some Aspects of the Mathematical Theory of Control Processes*, Rand Corporation, R-313, 1958.

LINEAR BINARY TWO-SAMPLE TESTS FOR SHIFT ALTERNATIVES

G. V. MARTYNOV

(*Translated by B. Seckler*)

0. A class of linear binary tests based on pairwise comparisons of observations of two samples is considered. Sign tests and Wilcoxon's test in particular belong to this class. On the basis of the notion of asymptotic relative efficiency (A.R.E.), linear binary tests are compared with Student's t-test for alternative shifts when the distributions of the sample observations have a density function belonging to $L_2(-\infty, \infty)$-space. In conclusion, a special case of a linear binary test is considered—the k-diagonal test—which is nearly as simple as the sign test and as efficient as the Wilcoxon test.

1. Let $\{A_n\}$ be a given sequence of sets of observations. For instance, $A_n = \{x_1, \cdots, x_n\}$ are sample observations of a random variable or $A_n = \{x_1, \cdots, x_{a_n}; y_1, \cdots, y_{b_n}\}$ are a_n and b_n observations of two samples. Suppose that the tests T and T' are based on the statistics $T_n = T(A_n)$ and $T'_n = T'(A_n)$ and are used to test hypotheses concerning some parameter θ. We shall compare these tests as to the A.R.E. of one with respect to the other.

Definition. With n indexing the size of the sets of observations, suppose that the test T_n (T'_n) is used with confidence level α_n (α'_n) such that $\lim_n \alpha_n = \lim_n \alpha'_n = \alpha$, where α is a given positive