The Annals of Statistics
1975, Vol. 3, No. 1, 1-14

A QUADRATIC MEASURE OF DEVIATION OF TWO-DIMENSIONAL DENSITY ESTIMATES AND A TEST OF INDEPENDENCE[1]

BY M. ROSENBLATT

University of California, San Diego

This paper considers estimates of multidimensional density functions based on a bounded and bandlimited weight function. The asymptotic behavior of quadratic functions of density function estimates useful in setting up a test of goodness of fit of the density function is determined. A test of independence is also given. The methods use a Poissonization of sample size. The estimates considered are appropriate if interested in estimating density functions or determining local deviations from a given density function.

1. Introduction. Let ${}_1X, {}_2X, \cdots, {}_nX, \cdots$ be independent and identically distributed random two vectors with continuous density function $f(x)$, $x = (x_1, x_2)$. We consider a class of estimates $f_n(x)$ of $f(x)$ determined by a bounded weight function w with finite support

$$
\begin{aligned}
(1) \qquad f_n(x) &= \frac{1}{nb(n)^2} \sum_{j=1}^{n} w\left(\frac{x - {}_jX}{b(n)}\right) \\
&= \int \frac{1}{b(n)^2} w\left(\frac{x-s}{b(n)}\right) dF_n(s) ,
\end{aligned}
$$

where F_n is the sample distribution function determined by ${}_jX, j = 1, \cdots, n$. Here $b(n)$ is a bandwidth such that $b(n) \downarrow 0$ and $nb(n)^2 \to \infty$ as $n \to \infty$.

The object of this paper is to describe the asymptotic behavior of quadratic functionals of density function estimates that are useful in setting up a test of goodness-of-fit of a density function and a test of independence. The methods are different in part from those used in [1] and involve a Poissonization of the sample size used because of the multidimensional context. Remarks will be made later in contrasting the type of test of goodness-of-fit obtained here with a typical example of such a test making use of the sample distribution function. The methods employed here can be used in the general multidimensional case; the two-dimensional case examined being typical. The statistics proposed here are plausible in situations in which one is specifically interested in estimating the density function or in local departures from a given density function (see [3] and [6]).

Received June 1973; revised February 1974.

[1] Research supported by the Office of Naval Research.

AMS 1970 *subject classifications.* Primary 62G05, 62H30; Secondary 65D10.

Key words and phrases. Multidimensional density estimates, weight function, bandwidth. quadratic measure, asymptotic distribution, test of goodness of fit, test of independence.

It is convenient at this point to state certain assumptions which we shall refer to as a1, a2, a3, a4:

a1. The weight function w is bounded and equal to zero outside the square $[\frac{1}{2}, \frac{1}{2}]^2$. Further

$$\int w(x)\, dx = 1 \,. \tag{2}$$

a2. The probability density function f is bounded and either positive on all of R^2, or else positive on $[0, 1]^2$ and zero outside the unit square. Also f is continuously differentiable up to second order with bounded derivatives in the interior of its domain of positivity.

a3. The weight function w is symmetric ($w(u) = w(-u)$) so that the first moments of w, w^2 are zero, and the matrix of second moments of w

$$\int x_j x_k w(x)\, dx \,, \qquad j, k = 1, 2 \tag{3}$$

is positive definite.

a4. The function a is a bounded integrable function whose set of discontinuities has two-dimensional Jordan content zero.

A number of the more important results of the paper are stated below. The proofs will be given later on.

THEOREM 1. *Under assumptions* a1, a2, a3, a4

$$b(n)^{-1}\{nb(n)^2 \int [f_n(x) - f(x)]^2 a(x)\, dx - \int f(x)a(x)\, dx \cdot \int w(u)^2\, du\} \tag{4}$$

is asymptotically normally distributed with mean zero and variance

$$2w^{(4)}(0) \int a(x)^2 f(x)^2\, dx \tag{5}$$

as $n \to \infty$ if $nb(n)^2 \to \infty$ and $b(n) = o(n^{-\frac{1}{5}})$.

THEOREM 2. *Let the components of the random variables ${}_jX$ be independent. Assume that $g_n(x_1)$, $h_n(x_2)$ are the estimates of the marginal density functions $g(x_1)$, $h(x_2)$ given in* (42) *and that $w(x) = w_1(x_1)w_2(x_2)$. Then under the conditions of Theorem* 1

$$b(n)^{-1}[nb(n)^2 \int [f_n(x) - g_n(x_1)h_n(x_2)]^2 a(x)\, dx - A(n)] \tag{6}$$

is asymptotically normally distibuted with mean zero and variance (5). *Here*

$$\begin{aligned} A(n) = \int f(x)a(x)\, dx &+ b(n) \int g(x_1)h(x_2)^2 a(x)\, dx[1 + \int w_1(u_1)^2 w_2(u_2)\, du] \\ &+ b(n) \int g(x_1)^2 h(x_2)a(x)\, dx[1 + \int w_1(u_1)w_2(u_2)^2\, du] \,. \end{aligned} \tag{7}$$

COROLLARY 1. *Suppose one considers the one-dimensional analogue of the result obtained in Theorem* 1 *with* a1′—a4′ *the corresponding one-dimensional assumptions. Then*

$$b(n)^{-\frac{1}{2}}\{nb(n) \int [f_n(x) - f(x)]^2 a(x)\, dx - \int f(x)a(x)\, dx \int w(u)^2\, du\} \tag{8}$$

is asymptotically normally distributed with mean zero and variance (5) *as $n \to \infty$ if $nb(n) \to \infty$ and $b(n) = o(n^{-\frac{2}{9}})$.*

Notice that Corollary 1 can be regarded as a substantial improvement on the Theorem stated in Section 1 of [1] for the limiting distribution of (8) since the assumptions in Corollary 1 on $b(n)$ are weaker. This suggests that the technique of Poissonization coupled with Lemmas 1 and 2 of this paper are a more natural technique for dealing with quadratic functionals than that of Skorohod imbedding.

2. Proofs. A number of lemmas that are required for the principal results are derived first. A number of lettered subheadings will be used in this section to indicate the primary topics.

LEMMA 1. *Let w be a bounded integrable weight function with first moments zero, $\int w(x)\,dx = 1$, and*

$$|\int (x_1^2 + x_2^2)w(x)\,dx| < \infty . \tag{9}$$

If assumptions a2, a4 *are also satisfied, then*

$$\begin{aligned} A &= nb(n)^2[\int \{(f_n(x) - f(x))^2 - (f_n(x) - Ef_n(x))^2\}a(x)\,dx] \\ &= o(b(n)) , \end{aligned} \tag{10}$$

if $b(n) = o(n^{-\frac{1}{3}})$.

Notice that

$$\begin{aligned} A &= nb(n)^2 \int [(Ef_n(x) - f(x))^2 + 2(Ef_n(x) - f(x))(f_n(x) - Ef_n(x))]a(x)\,dx \\ &= A_1 + A_2 . \end{aligned} \tag{11}$$

The first expression

$$A_1 = O(nb(n)^6)$$

under the assumptions on w and a2, a4. Also $E|A_2| \leqq \sigma(A_2)$. Since

$$\begin{aligned} \operatorname{Cov}(f_n(x), f_n(x')) &= \frac{1}{nb(n)^4} \operatorname{Cov}\left(w\left(\frac{x - X}{b(n)}\right), w\left(\frac{x' - X}{b(n)}\right)\right) \\ &= \frac{1}{nb(n)^4}\left\{\int w\left(\frac{x - u}{b(n)}\right) w\left(\frac{x' - u}{b(n)}\right) f(u)\,du \right. \\ &\qquad \left. - \int w\left(\frac{x - u}{b(n)}\right) f(u)\,du \int w\left(\frac{x' - v}{b(n)}\right) f(v)\,dv\right\} , \end{aligned}$$

it follows that

$$\begin{aligned} \sigma(A_2) &\leqq 2nb(n)^2 \left\{\int \left|\int w\left(\frac{x - u}{b(n)}\right)[Ef_n(x) - f(x)]a(x)\,dx\right|^2 f(u)\,du\, n^{-1}b(n)^{-4}\right\}^{\frac{1}{2}} \\ &\leqq kn^{\frac{1}{2}}b(n)^4 , \end{aligned} \tag{12}$$

where k is constant. It is clear that $A = o(b(n))$ if $b(n) = o(n^{-\frac{1}{3}})$.

The following remarks indicate what the corresponding result ought to be in the k-dimensional case. Then we have $nb(n)^k$ as a multiplicative factor instead of $nb(n)^2$. The corresponding terms A_1, A_2 are such that

$$A_1 = O(nb(n)^{4+k})$$

and

$$\sigma(A_2) \leqq Cn^{\frac{1}{2}}b(n)^{2+k}$$

with C a constant. Thus

$$A = o\left(b(n)^{k/2}\right)$$

if

$$b(n) = o\left(n^{-1/(4+k/2)}\right).$$

A. *Poissonization.* In order to complete a set of estimates required in the proof of Theorem 1, it is convenient to introduce a Poissonized sample size N. Let N be a Poisson random variable with mean n independent of ${}_1X, {}_2X, \cdots$. Set

$$f_n^*(x) = \frac{1}{nb(n)^2} \sum_{j=1}^{N} w\left(\frac{x - {}_jX}{b(n)}\right). \tag{13}$$

Then

$$f_n(x) - f_n^*(x) = \frac{1}{nb(n)^2} \sum{}'^{n}_{j=N} w\left(\frac{x - {}_jX}{b(n)}\right), \tag{14}$$

where by $\sum'^{n}_{j=N}$ we mean $\sum_{j=1}^{n} - \sum_{j=1}^{N}$. Since $Ef_n(x) = Ef_n^*(x)$

$$E|f_n(x) - f_n^*(x)|^2 = \frac{1}{n^2b(n)^4} E|N - n|\sigma^2\left(w\left(\frac{x - X}{b(n)}\right)\right).$$

Now $E|N - n| \leqq (E|N - n|^2)^{\frac{1}{2}} \leqq n^{\frac{1}{2}}$ and

$$\sigma^2\left(w\left(\frac{x - X}{b(n)}\right)\right) = b(n)^2 f(x) \int w^2(u)\, du + O(b(n)^4),$$

if a2 holds. Thus

$$E|f_n(x) - f_n^*(x)|^2 \leqq \frac{K}{n^{\frac{3}{2}}b(n)^2}, \tag{15}$$

where K is a constant.

LEMMA 2. *Under the assumptions of Lemma* 1

$$\begin{aligned} B &= nb(n)^2 \int \{(f_n(x) - Ef_n(x))^2 - (f_n^*(x) - Ef_n^*(x))^2\} a(x)\, dx \\ &= o(b(n)), \end{aligned} \tag{16}$$

if $nb(n)^2 \to \infty$ *and* $b(n) \downarrow 0$ *as* $n \to \infty$.

Now

$$\begin{aligned} B = nb(n)^2 \int \{&(f_n(x) - f_n^*(x))^2 \\ &+ 2(f_n^*(x) - Ef_n^*(x))(f_n(x) - f_n^*(x))\} a(x)\, dx \\ = B_1 + B_2&. \end{aligned} \tag{17}$$

From (15) it follows that

$$EB_1 \leqq K'n^{-\frac{1}{2}} \tag{18}$$

with K' a constant. To estimate B_2 we look at

$$\text{(19)} \qquad C = E(f_n^*(x) - Ef_n^*(x))(f_n(x) - f_n^*(x))(f_n^*(x') - Ef_n^*(x')) \times (f_n(x') - f_n^*(x')) .$$

Notice that

$$f_n^*(x) - Ef_n^*(x) = n^{-1}b(n)^{-2} \sum_{j=1}^{N} \alpha_j(x)$$

with

$$\alpha_j(x) = w\left(\frac{x - {}_jX}{b(n)}\right) - Ew\left(\frac{x - {}_jX}{b(n)}\right),$$

and

$$f_n(x) - f_n^*(x) = n^{-1}b(n)^{-2} \sum\nolimits'^{n}_{j=N} \beta_j(x)$$

with

$$\beta_j(x) = w\left(\frac{x - {}_jX}{b(n)}\right).$$

Then (19) can be written as

$$\text{(20)} \qquad n^{-4}b(n)^{-8}E\{\sum_{j_1=1}^{N} \alpha_{j_1}(x) \sum\nolimits'^{n}_{j_2=N} \beta_{j_2}(x) \sum_{j_3=1}^{N} \alpha_{j_3}(x') \sum\nolimits'^{n}_{j_4=N} \beta_{j_4}(x')\} .$$

We have to distinguish between (a) $N = m > n$ and (b) $N = m \leqq n$. In case (b) the contribution to the expectation in (20) is

$$\text{(21)} \qquad \sum_{m=0}^{n} \frac{n^m e^{-n}}{m!} mE[\alpha(x)\alpha(x')] \times \{(n-m)E[\beta(x)\beta(x')] + (n-m)(n-m-1)E\beta(x)E\beta(x')\} .$$

In case (a) the contribution to the expectation in (20) is

$$\text{(22)} \qquad \begin{aligned} \sum_{m=n+1}^{\infty} \frac{n^m e^{-n}}{m!} \{ & nE[\alpha(x)\alpha(x')] \\ & \times ((m-n)E[\beta(x)\beta(x')] + (m-n)(m-n-1)E\beta(x)E\beta(x')) \\ & + (m-n)E[\alpha(x)\alpha(x')\beta(x)\beta(x')] \\ & + (m-n)E[\alpha(x)\alpha(x')] \\ & \times ((m-n-1)E[\beta(x)\beta(x')] \\ & + (m-n-1)(m-n-2)E\beta(x)E\beta(x')) \\ & + (E[\alpha(x)\beta(x)]E[\alpha(x')\beta(x')] + E[\alpha(x)\beta(x')]E[\alpha(x')\beta(x)]) \\ & \times (m-n)(m-n-1) \\ & + (E[\alpha(x)\alpha(x')\beta(x)]E\beta(x') + E[\alpha(x)\alpha(x')\beta(x')]E\beta(x)) \\ & \times (m-n)(m-n-1)\} . \end{aligned}$$

Let

$$\begin{aligned} x^+ &= x \qquad \text{if} \quad x \geqq 0 \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

The following inequalities

$$
\begin{aligned}
Em(n-m)^+ &= nE(n-m)^+ - E[(n-m)^+]^2 \leqq n^{\frac{3}{2}} , \\
Em[(n-m)^+]^2 &= nE[(n-m)^+]^2 - E[(n-m)^+]^3 \leqq n^2 , \\
En(m-n)^+ &= nE(m-n)^+ \leqq n^{\frac{3}{2}} , \\
En[(m-n)^+]^2 &= nE[(m-n)^+]^2 \leqq n^2 , \\
E[(m-n)^+]^3 &\leqq \{E(m-n)^4\}^{\frac{3}{4}} \leqq (4n^2)^{\frac{3}{4}} \leqq 4n^{\frac{3}{2}}
\end{aligned}
$$

follow from the fact that the characteristic function of a Poisson random variable with mean n is

$$\exp\{n(e^{it}-1) - itn\} = \exp\left\{-\frac{t^2}{2}n + \frac{(it)^3}{3!}n + \cdots\right\} .$$

Now

$$EB_2^2 = 4 \int Ca(x)a(x')\, dx\, dx'\, n^2 b(n)^4 , \tag{23}$$

where C is given by (19). Notice that

$$
\begin{aligned}
E\beta(x) &= Ew\left(\frac{x-X}{b(n)}\right) = \int w\left(\frac{x-u}{b(n)}\right) f(u)\, du \\
&= \int w(z) f(x - b(n)z)\, dz\, b(n)^2 \cong f(x) \int w(z)\, dz\, b(n)^2 ,
\end{aligned} \tag{24}
$$

$$
\begin{aligned}
E\beta(x)\beta(x') &= Ew\left(\frac{x-X}{b(n)}\right) w\left(\frac{x'-X}{b(n)}\right) \\
&= \int w\left(\frac{x-u}{b(n)}\right) w\left(\frac{x'-u}{b(n)}\right) f(u)\, du \\
&= \int w(z)\, w\left(\frac{x'-x}{b(n)} + z\right) f(x - b(n)z)\, dz\, b(n)^2 .
\end{aligned} \tag{25}
$$

Then, using (24) and (25) one finds that

$$
\begin{aligned}
EB_2^2 \leqq Kn^{-2}b(n)^{-4}\{ & n^{\frac{3}{2}}b(n)^6 + n^2 b(n)^8 + n^{\frac{1}{2}}b(n)^4 + nb(n)^6 + n^{\frac{3}{2}}b(n)^8 \\
& + nb(n)^4 + nb(n)^6 + nb(n)^6 + nb(n)^6\} ,
\end{aligned} \tag{26}
$$

and this implies that

$$E|B_2| = o(b(n)) ,$$

if $nb(n)^2 \to \infty$ and $b(n) \downarrow 0$ as $n \to \infty$. The conclusion of Lemma 2 then follows.

In the k-dimensional case we would again have $nb(n)^k$ instead of $nb(n)^2$ as a multiplicative factor in B. Since

$$E|f_n(x) - f_n^*(x)|^2 \leqq C/(n^{\frac{3}{2}}b(n)^k) ,$$

then

$$B = o(b(n)^{k/2}) ,$$

if $nb(n)^k \to \infty$ and $b(n) \downarrow 0$ as $n \to \infty$.

B. *Moments of a Poisson process on the plane.* Before continuing with the proof of Theorem 1 it is helpful to make a simple remark about moments of an estimate of the distribution function with Poisson sample size N. If F_N is the sample distribution function with Poisson sample size N, let $F_n^* = (N/n)F_N$. The characteristic function of a Poisson random variable with mean λ is

$$\exp\{\lambda(e^{it} - 1) - it\lambda\} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} [e^{it} - 1 - it]^k . \tag{27}$$

The advantage of considering a Poisson sample size N lies in the fact that nF_n^* is then a Poisson process on the plane so that the number of points in fixed disjoint sets is independent, and the mean number of points in for example a rectangle is n times the increment of F over that rectangle. This remark and formula (27) imply that

$$n^k E(d(F_n^* - F))^{2k} = \frac{(2k)!}{k!\,2^k} (dF)^k + \sum_{j=1}^{k-1} a_{j,n}^{(2k)} (dF)^j , \tag{28}$$

$$n^{(2k+1)/2} E(d(F_n^* - F))^{2k+1} = \sum_{j=1}^{k} a_{j,n}^{(2k+1)} (dF)^j ,$$

where

$$a_{j,n}^{(s)} = O(n^{-((s/2)-j)}) \tag{29}$$

as for each fixed j, s with $j = 1, \cdots, (s/2) - 1$ and (u) is the smallest integer greater than or equal to u.

Note that

$$f_n^*(x) - Ef_n^*(x) = \int \frac{1}{b(n)^2} w\left(\frac{x-u}{b(n)}\right) d(F_n^*(u) - F(u)) . \tag{30}$$

Lemmas 1 and 2 will tell us that if we let $b(n) \downarrow 0$ at the proper rate,

$$S_n = nb(n)^2 \int [f_n^*(x) - Ef_n^*(x)]^2 a(x)\, dx \tag{31}$$

can be considered instead of

$$nb(n)^2 \int [f_n(x) - f(x)]^2 a(x)\, dx ,$$

with small error as $n \to \infty$, $b(n) \downarrow 0$.

C. *Asymptotic normality*. We now continue with the proof of Theorem 1. Note that

$$S_n = \sum_{j,k} U_{j,k}(n) , \tag{32}$$

where

$$U_{j,k}(n) = \int_{jb(n)}^{(j+1)b(n)} \int_{kb(n)}^{(k+1)b(n)} \left[\frac{n^{\frac{1}{2}}}{b(n)} \int w\left(\frac{x_1 - u_1}{b(n)}, \frac{x_2 - u_2}{b(n)}\right) d(F_n^* - F)\right]^2 \times a(x_1, x_2)\, dx_1\, dx_2 . \tag{33}$$

The random variables $U_{j,k}(n)$ would be called 2×2 dependent, that is, $\{U_{j,k}(n)\}$ and $\{U_{j',k'}(n)\}$ are independent if $|j - j'| \geqq 2$ or $|k - k'| \geqq 2$ for all pairs (j, k) and (j', k'). This follows from the assumption that w is zero outside $[-\frac{1}{2}, \frac{1}{2}]^2$ and the fact that nF_n^* is a Poisson process on the plane. By using this 2×2

dependence, a central limit theorem is proven for S_n just as for a k-step dependent process with a one-dimensional parameter.

The mean

$$(34)\qquad ES_n = \int \frac{1}{b(n)^2} \int w\left(\frac{x-u}{b(n)}\right)^2 f(u)\,du\,a(x)\,dx$$
$$= \int f(x)a(x)\,dx \cdot \int w(u)^2\,du + O(b(n)^2)\,,$$

if f is bounded and continuously differentiable up to 2nd order with bounded derivatives and the first moments of w^2 are zero.

Let

$$(35)\qquad V_{j,k}(n) = \int_{\Delta_j}^{\Delta_j'} \int_{\Delta_k}^{\Delta_k'} \left[\frac{n^{\frac{1}{2}}}{b(n)} \int w\left(\frac{x-u}{b(n)}\right) d(F_n^*(u) - F(u))\right]^2 a(x)\,dx_1\,dx_2\,,$$

where $x = (x_1, x_2)$ and

$$\Delta_j = (j+1)b(n) + j\Delta \qquad \Delta_j' = (j+1)(b(n)+\Delta)$$

with $b(n) = o(\Delta)$ where $\Delta = \Delta(n) \downarrow 0$ as $n \to \infty$. First notice that the random variables $V_{j,k}$ are independent. They play the role of the big blocks in a usual argument to get a central limit theorem. These big blocks are separated by thin strips of thickness $b(n)$.

Notice that

$$(36)\qquad EV_{j,k}(n) = \int_{\Delta_j}^{\Delta_j'} \int_{\Delta_k}^{\Delta_k'} \frac{1}{b(n)^2} \int w\left(\frac{x-u}{b(n)}\right)^2 f(u)\,du\,a(x)\,dx_1\,dx_2\,,$$

so that

$$(37)\qquad E\left|\sum_{j,k} V_{j,k}(n) - S_n\right| = O(b(n)\Delta)\,.$$

Let $I(j,k) = (\Delta_j, \Delta_j') \times (\Delta_k, \Delta_k')$. Formula (28) and the character of the process nF_n^* imply that

$$n^2 \operatorname{Cov}\{d(F_n^*(u) - F(u))\,d(F_n^*(u') - F(u')),$$
$$d(F_n^*(v) - F(v))\,d(F_n^*(v') - F(v'))\}$$
$$= [\delta(u-v)\delta(u'-v') + \delta(u-v')\delta(u'-v)]\,dF(u)\,dF(u')$$
$$+ \delta(u-u')\delta(v-v')\delta(u-v)\left\{(dF(u))^2 + \frac{1}{n}\,dF(u)\right\};$$

$$\sigma^2(V_{j,k}(n)) = \frac{1}{b(n)^4} \int_{I(j,k)} a(x) \int_{I(j,k)} a(x') \left\{2\left[\int w\left(\frac{x-u}{b(n)}\right) w\left(\frac{x'-u}{b(n)}\right) f(u)\,du\right]^2\right.$$
$$\left. + \frac{1}{n} \int w\left(\frac{x-u}{b(n)}\right)^2 w\left(\frac{x'-u}{b(n)}\right)^2 f(u)\,du\right\} dx\,dx'$$
$$\cong w^{(4)}(0) \int_{I(j,k)} a^2(x)\left[2b(n)^2 f(x)^2 + \frac{1}{n} f(x)\right] dx\,,$$

and this implies that

$$(38)\qquad \sigma^2\left(\sum V_{j,k}(n)\right) \cong 2b(n)^2 w^{(4)}(0) \int a(x)^2 f(x)^2\,dx\,.$$

Further

$$
\begin{aligned}
(39)\qquad & E|V_{j,k}(n) - EV_{j,k}(n)|^4 \\
&= 3\sigma^4(V_{j,k}(n)) \\
&\quad + 48(\textstyle\int_{I(j,k)})^4 \prod_{l=1}^4 dx^{(l)}(\int)^4 \prod_{l=1}^4 du^{(l)} \prod_{l=1}^4 a(x^{(l)}) \prod_{l=1}^4 f(u^{(l)}) b(n)^{-8} \\
&\quad \times w\left(\frac{x^{(1)} - u^{(1)}}{b(n)}\right) w\left(\frac{x^{(2)} - u^{(1)}}{b(n)}\right) w\left(\frac{x^{(1)} - u^{(2)}}{b(n)}\right) w\left(\frac{x^{(3)} - u^{(2)}}{b(n)}\right) \\
&\quad \times w\left(\frac{x^{(2)} - u^{(3)}}{b(n)}\right) w\left(\frac{x^{(4)} - u^{(3)}}{b(n)}\right) w\left(\frac{x^{(3)} - u^{(4)}}{b(n)}\right) w\left(\frac{x^{(4)} - u^{(4)}}{b(n)}\right) \\
&\quad + O(b(n)^6\Delta^2) \\
&= O(b(n)^4\Delta^4) .
\end{aligned}
$$

Let $S_n(R)$ be expression (31) with the integral only extended over the rectangle $R = [-r, r]^2$. If r is large the variance of $S_n - S_n(R)$ is small. By Liapunov's version of the central limit theorem, $\sum V_{j,k}$ extended over summands $V_{j,k}$ arising from integrals lying in R (r fixed) is asymptotically normal with mean $ES_n(R)$ and variance $\sigma^2(S_n(R))$ as $n \to \infty$ if $b(n) = o(\Delta)$ with $\Delta = \Delta(n) \downarrow 0$ as $n \to \infty$. Then $S_n(R)$ is asymptotically normal with the same mean and variance. A standard approximation argument shows that S_n is asymptotically normal with mean (34) and variance (38). The proof of Theorem 1 is complete.

D. *The one- and k-dimensional cases.* Corollary 1 follows by noting that in the one-dimensional case under parallel assumptions

$$A = o(b(n)^{\frac{1}{2}}) ,$$

if $b(n) = o(n^{-\frac{2}{9}})$ and

$$B = o(b(n)^{\frac{1}{2}}) ,$$

if $nb(n) \to \infty$ and $b(n) \to 0$ as $n \to \infty$.

In the k-dimensional case under corresponding assumptions

$$A = o(b(n)^{k/2}) ,$$

if $b(n) = o(n^{-2/(8+k)})$ and

$$B = o(b(n)^{k/2}) ,$$

if $nb(n)^k \to \infty$ and $b(n) \to 0$ as $n \to \infty$. The k-dimensional version of Corollary 1 can then easily be written out as follows:

COROLLARY 1′. *Consider the k-dimensional analogue of Theorem* 1 *with* A1″—A4″ *the corresponding k-dimensional assumptions. Then*

$$b(n)^{-k/2}\{nb(n)^k \textstyle\int [f_n(x) - f(x)]^2 a(x)\, dx - \int f(x)a(x)\, dx \int w(u)^2\, du\}$$

is asymptotically normally distributed with mean zero and variance (5) *as* $n \to \infty$ *if* $nb(n)^k \to \infty$ *and* $b(n) = o(n^{-2/(8+k)})$.

E. *Independent components.* In the case the component random variables

X_1, X_2 of X are independent, it is interesting to look at

$$f_n(x_1, x_2) - g_n(x_1)h_n(x_2)\,, \tag{40}$$

where

$$\begin{aligned} &f(x_1, x_2) = g(x_1)h(x_2) \\ &w(x_1, x_2) = w_1(x_1)w_2(x_2) \\ &\int g(x_1)\,dx_1 = \int h(x_2)\,dx_2 = \int w_1(x_1)\,dx_1 = \int w_2(x_2)\,dx_2 = 1\,, \end{aligned} \tag{41}$$

and

$$\begin{aligned} g_n(x_1) &= \frac{1}{nb(n)} \sum_{j=1}^{n} w_1\left(\frac{x_1 - {}_jX_1}{b(n)}\right) \\ h_n(x_2) &= \frac{1}{nb(n)} \sum_{j=1}^{n} w_2\left(\frac{x_2 - {}_jX_2}{b(n)}\right). \end{aligned} \tag{42}$$

Notice that

$$\begin{aligned} n^{\frac{1}{2}}b(n)&[f_n(x_1, x_2) - g_n(x_1)h_n(x_2)] \\ &= n^{\frac{1}{2}}b(n)[f_n(x_1, x_2) - Ef_n(x_1, x_2)] \\ &\quad - \frac{1}{n^{\frac{1}{2}}}(nb(n))^{\frac{1}{2}}[g_n(x_1) - Eg_n(x_1)](nb(n))^{\frac{1}{2}}[h_n(x_2) - Eh_n(x_2)] \\ &\quad - (b(n))^{\frac{1}{2}}(nb(n))^{\frac{1}{2}}[g_n(x_1) - Eg_n(x_1)]Eh_n(x_2) \\ &\quad - (b(n))^{\frac{1}{2}}(nb(n))^{\frac{1}{2}}[h_n(x_2) - Eh_n(x_2)]Eg_n(x_1) \\ &= I_1 + I_2 + I_3 + I_4\,, \end{aligned} \tag{43}$$

since

$$Ef_n(x_1, x_2) = Eg_n(x_1)Eh_n(x_2)\,.$$

The proof of Theorem 2 is now given.

We wish to look at

$$\begin{aligned} nb(n)^2 \int [f_n(x_1, x_2) &- g_n(x_1)h_n(x_2)]^2 a(x)\,dx_1\,dx_2 \\ &= \int I_1^2 a\,dx + \int I_2^2 a\,dx + \int I_3^2 a\,dx + \int I_4^2 a\,dx \\ &\quad + 2\int I_1 I_2 a\,dx + 2\int I_1 I_3 a\,dx + 2\int I_1 I_4 a\,dx \\ &\quad + 2\int I_2 I_3 a\,dx + 2\int I_2 I_4 a\,dx + 2\int I_3 I_4 a\,dx\,. \end{aligned} \tag{44}$$

The contribution from the first term on the right we already know. The second term, on taking its expectation is seen to be $O(1/n) = o(b(n))$ and so can be disregarded. The third term looks like

$$nb(n)^2 \int [g_n(x_1) - Eg_n(x_1)]^2 \beta(x_1)\,dx_1\,,$$

where

$$\beta(x_1) = \int h(x_2)^2 a(x_1, x_2)\,dx_2\,.$$

Thus the third term is to the first order

$$b(n) \int g(x_1)h(x_2)^2 a(x_1, x_2)\,dx_1\,dx_2\,.$$

Similarly the fourth term is to the first order

$$b(n) \int g(x_1)^2 h(x_2) a(x_1, x_2)\, dx_1\, dx_2 .$$

Using the Schwarz inequality one can see that the fifth term is

$$O\left(\frac{1}{n^{\frac{1}{2}}}\right) = o(b(n)) ,$$

if $nb(n)^2 \to \infty$ as $n \to \infty$. We now look at the sixth term. Note that

$$\begin{aligned} E\{[f_n(x) - Ef_n(x)][g_n(x_1) - Eg_n(x_1)]\} \\ &\cong n^{-1}b(n)^{-3} \int w_1\left(\frac{x_1 - u_1}{b(u)}\right)^2 w_2\left(\frac{x_2 - u_2}{b(n)}\right) f(u_1, u_2)\, du_1\, du_2 \\ &\cong n^{-1}b(n) f(x_1, x_2) \int w_1(u_1)^2 w_2(u_2)\, du_1\, du_2 . \end{aligned}$$

This implies that the mean of this term is to the first order

$$b(n) \int w_1(x_1)^2 w_2(x_2)\, dx \int f(x) h(x_2) a(x)\, dx .$$

The variance of the sixth term is

$$\begin{aligned} \frac{1}{n^2 b(n)^2} \Bigg\{ &\int h(x_2) h(x_2') a(x) a(x')\, dx\, dx' \\ &\times \left(nE\left[w\left(\frac{x - X}{b(n)}\right) - E \cdots \right]\left[w_1\left(\frac{x_1 - X_1}{b(n)}\right) - E \cdots \right] \right. \\ &\times \left[w\left(\frac{x' - X}{b(n)}\right) - E \cdots \right]\left[w_1\left(\frac{x_1' - X_1'}{b(n)}\right) - E \cdots \right] \\ &+ (n^2 - n)\left(E\left[w\left(\frac{x - X}{b(n)}\right) - E \cdots \right]\left[w_1\left(\frac{x_1' - X_1}{b(n)}\right) - E \cdots \right] \right. \\ &\times E\left[w\left(\frac{x' - X}{b(n)}\right) - E \cdots \right]\left[w_1\left(\frac{x_1 - X_1}{b(n)}\right) - E \cdots \right] \\ &+ E\left[w\left(\frac{x - X}{b(n)}\right) - E \cdots \right]\left[w\left(\frac{x' - X}{b(n)}\right) - E \cdots \right] \\ &\left.\left. \times E\left[w_1\left(\frac{x_1' - X_1}{b(n)}\right) - E \cdots \right]\left[w_1\left(\frac{x_1 - X_1}{b(n)}\right) - E \cdots \right] \right)\right) \Bigg\} \end{aligned} \tag{45}$$

where an expression of the form $\alpha(X) - E \cdots$ denotes $\alpha(X) - E\alpha(X)$. The variance (45) can be shown to be $O(b(n)^3)$. The mean of the sixth term must be taken into account since it is $O(b(n))$, but the fluctuation about the mean can be neglected since it is $O(b(n)^{\frac{3}{2}})$. The analysis for the seventh term proceeds just as does that for the sixth term. The mean to the first order is

$$b(n) \int w_1(x_1) w_2(x_2)^2\, dx_1\, dx_2 \int f(x) g(x_1) a(x)\, dx$$

and the variance is $O(b(n)^{\frac{3}{2}})$. The Schwarz inequality can be used to show that the eighth and ninth terms can be neglected. The mean of the last term is zero. Also the variance of this last term is small enough so that it can be neglected.

Thus, if

(46) $$A(n) = \int f(x)a(x)\,dx + b(n)\int g(x_1)h(x_2)^2 a(x)\,dx[1 + \int w_1(u_1)^2 w_2(u_2)\,du] + b(n)\int g(x_1)^2 h(x_2)a(x)\,dx[1 + \int w_1(u_1)w_2(u_2)^2\,du]\,,$$

then

$$b(n)^{-1}[nb(n)^2 \int [f_n(x) - g_n(x_1)h_n(x_2)]^2 a(x)\,dx - A(n)]$$

is asymptotically normally distributed with mean zero and variance

(47) $$2w^{(4)}(0)\int a(u)^2 f(u)^2\,du\,,$$

given that $nb(n)^2 \to \infty$ and $b(n) = o(n^{-\frac{1}{5}})$ as $n \to \infty$. The proof of Theorem 2 is complete.

In the paper [1] it was noted that

$$\sup_{0 \leq x_1 \leq 1} (nb(n))^{\frac{1}{2}} |g_n(x_1) - g(x_1)| = O((\log n)^{\frac{1}{2}})\,.$$

By using this result we obtain the following corollary.

COROLLARY 2. *Let the components of the random variables ${}_jX$ be independent. Further, assume that the marginal densities are positive on* $[0, 1]$ *and zero outside* $[0, 1]$. *Then under the assumptions of Theorem* 2

(47) $$b(n)^{-1}\left[nb(n)^2 \int_{(b(n))^{\frac{1}{2}}}^{1-(b(n))^{\frac{1}{2}}} \int_{(b(n))^{\frac{1}{2}}}^{1-(b(n))^{\frac{1}{2}}} \frac{[f_n(x) - g_n(x_1)h_n(x_2)]^2}{g_n(x_1)h_n(x_2)}\,dx - A(n)\right]$$

is asymptotically normally distributed with mean zero and variance $2w^{(4)}(0)$ *if* $b(n) = n^{-\delta}$ *with* $\frac{1}{3} > \delta > \frac{1}{5}$. *Here*

(48) $$A(n) = 1 + b(n)[2 + \int w_1(u_1)^2\,du_1 + \int w_2(u_2)^2\,du_2]\,.$$

To make local power computations in the context of Theorem 1 we consider the behavior of the statistic for a sequence of alternatives to $f_0(x)$ of the form

(49) $$r_n(x) = f_0(x) + \gamma_n \eta(x) + o(\gamma_n)\,,$$

where the r_n satisfies a2 uniformly in n, $\gamma_n \downarrow 0$ at a rate to be specified and $o(\gamma_n)$ is uniform in x. The function η is assumed continuous.

F. *Asymptotic power.* The following results on asymptotic power are obtained just as in [1].

THEOREM 3. *Let* ${}_1X, {}_2X, \cdots, {}_nX$ *be independent and identically distributed with common density function* $r_n(x)$. *Let the assumptions* a1, a3, a4 *be satisfied. Set* $T_n = nb(n)^2 \int [f_n(x) - f_0(x)]a(x)\,dx$. *Then if* $b(n) = n^{-\delta}$ *with* $\frac{1}{3} > \delta > \frac{1}{5}$ *where* $\gamma_n = n^{-\frac{1}{2}}b(n)^{-\frac{1}{2}}$, *then it follows that*

(50) $$b(n)^{-1}[T_n - \int f_0(x)a(x)\,dx \int w(z)^2\,dz]$$

is asymptotically normally distributed with mean $\int \eta(x)^2 a(x)\,dx$ *and variance* $2w^{(4)}(0)\int a^2(x)f_0^2(x)\,dx$.

Similar computations for a sequence of alternatives of the form (49) with

$f_0(x) = g_0(x_1)h_0(x_2)$ would allow us to make local power computations for the statistic used to test independence in the corollary.

3. A comparison of tests based on estimates of the distribution function and density function. In the multidimensional case one could just as well test independence by using a statistic based on the sample distribution as in [2] rather than one based on the density function. It should be noted, however, that the limiting distribution obtained in the case of (47) is the normal distribution, rather than a distribution whose properties are not familiar as in [2]. Nonetheless, the local power computations of the type based on alternatives like (49) would appear to indicate that tests based on the sample distribution function are more powerful than those based on the density. There are other types of local alternatives for which tests based on density estimates are more powerful that tests based on the sample distribution function. So as to keep the argument simple notationally we shall illustrate this by an example in the one-dimensional case. Similar examples can be constructed in the multidimensional case. Let us consider the question of a test of goodness-of-fit. As is well known, the local alternatives when using the sample distribution function as a basis for a test would differ from the null hypothesis by $O(1/n^{\frac{1}{2}})$ where n is the sample size. We shall look at

$$T_n = \int [f_n(x) - f(x)]^2 a(x)\, dx\, nb(n)$$

as the statistic based on a density function estimate with the density

$$g_n(x) = f(x) + \eta_n(x)\,,$$

where $\eta_n(x) \to 0$ in a manner to be prescribed later on as $n \to \infty$. Assume that conditions a 1′—a 4′ referred to in Corollary 1 are satisfied. The type of local deviation $\eta_n(x)$ is assumed to be of the form

$$\eta_n(x) = \alpha_n u\left(\frac{x-c}{\gamma_n}\right),$$

where both α_n, $\gamma_n \downarrow 0$ at appropriate rates to be specified as $n \to \infty$, and u is continuously differentiable up to second order and band limited with $b(n) = o(\gamma_n)$, $b(n) \downarrow 0$ as $n \to \infty$.

One can show that

$$b(n)^{-\frac{1}{2}}(T_n - (\int f(x)a(x)\, dx) \int w^2(z)\, dz)$$

is asymptotically normal with mean

$$a(c) \int u(x)^2\, dx\,,$$

and variance

$$2w^{(4)}(0) \int a(x)^2 f^2(x)\, dx\,,$$

as $n \to \infty$ if for example

$$\alpha_n = \frac{1}{n^{\frac{1}{2}} b(n)^{\frac{2}{3}}}\,, \qquad \gamma_n = b(n)^{\frac{5}{6}}\,,$$

with $b(n) = n^{-\frac{1}{6}}$. Notice that the magnitude of the indefinite integral of $\eta_n(x)$ is of the order

$$\alpha_n \gamma_n = \frac{b(n)^{\frac{1}{2}}}{n^{\frac{1}{2}}} = n^{-\frac{5}{12}},$$

so that we have greater power against such an alternative than in the case of a test based on the deviation between the sample and true distribution function. Notice that the deviation $\eta_n(x)$ is local not only in that it is close to zero but also since it is highly centered about the point c. Of course, one could also have an alternative which is a finite sum of such functions $\eta_n(x)$ centered about different points.

I am indebted to Peter Bickel, who initially suggested Poissonization in discussions and computations leading up to our joint paper [1] as a possible technique. His comments and suggestions have been very helpful in the writing of this paper.

It is plausible though not absolutely clear that one might be able to get similar results for other density estimates of types mentioned in [4] and [5]. Also the more extended results of this paper are based on the use of weight functions with bounded support. It is not apparent how much additional labor would be required to prove them for a broader class of weight functions. In obtaining the results the estimates of Lemmas 1 and 2 were essential.

REFERENCES

[1] Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

[2] Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32** 485–498.

[3] Gibson, C. H. and Masiello, P. J. (1972). Observations of the variability of dissipation rates of turbulent velocity and temperature fields. *Statistical Models and Turbulence*, 427–453.

[4] Rosenblatt, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.

[5] Tarter, M. E. and Kronmal, R. A. (1970). On multivariate density estimates based on orthogonal expansions. *Ann. Math. Statist.* **41** 718–722.

[6] Van Atta, C. W. and Park, J. (1972). Statistical self-similarity and inertial subrange turbulence. *Statistical Models and Turbulence*, 402–426.

Department of Mathematics
University of California
La Jolla, California 92037