The Annals of Probability
1976, Vol. 4, No. 6, 1009-1015

ON THE MAXIMAL DEVIATION OF k-DIMENSIONAL DENSITY ESTIMATES

BY M. ROSENBLATT[1]

University of California, San Diego

Probability density estimates are generated by a kernel or weight function. Limit theorems are obtained for the maximum of the normalized deviation of the estimate from its expected value. The results are, in part, an extension to the k-dimensional case ($k > 1$) of those obtained by P. Bickel and M. Rosenblatt (*Ann. Statist.* 1973, 1071-1095) one dimensionally.

1. Introduction. Let $X_1, X_2, \cdots, X_n, \cdots$ be independent and identically distributed random k-vectors with continuous density function $f(x)$. We consider probability density estimators $f_n(x)$ determined by a weight function w of bounded support

$$(1)\qquad f_n(x) = \frac{1}{nb(n)^k} \sum_{j=1}^n w\left(\frac{x - X_j}{b(n)}\right) = \int \frac{1}{b(n)^k} w\left(\frac{x-s}{b(n)}\right) dF_n(s) .$$

In formula (1) F_n is the sample distribution function and $b(n)$ is a bandwidth that tends to zero as $n \to \infty$ and is such that $nb(n)^k \to \infty$. Components of a vector will be given by left indexing, that is, $x = ({}_jx, j = 1, \cdots, k)$. The asymptotic distribution of

$$(2)\qquad \tilde{M}_n = \max_{0 \leqq {}_jx \leqq 1; j=1,\cdots,k} n^{\frac{1}{2}} b(n)^{k/2} f(x)^{-\frac{1}{2}} |f_n(x) - Ef_n(x)|$$

as $n \to \infty$ will be determined under appropriate conditions on f and the weight function w.

There are a number of results that are of particular interest and will be given here for reference later on in the paper. The first of these is a remarkable result due to Komlós, Major and Tusnády [4]:

THEOREM A. *Let $X_1, X_2, \cdots, X_n, \cdots$ be independent random variables uniformly distributed on* [0, 1]. *A sequence of Brownian bridges $B_n(x)$ can be constructed so that*

$$(3)\qquad \sup_{0 \leqq x \leqq 1} |n^{\frac{1}{2}}(F_n(x) - x) - B_n(x)| = O(n^{-\frac{1}{2}} \log n)$$

almost surely.

This appears to be a "best" result. A one-dimensional Brownian bridge

$$B(x) = W(x) - xW(1) , \qquad 0 \leqq x \leqq 1 ,$$

Received September 10, 1975.

[1] The research was supported in part by the Office of Naval Research.

AMS 1970 *subject classifications.* Primary 60F10; Secondary 62G05.

Key words and phrases. k-dimensional density estimates, maximal deviation, Gaussian random field.

where $W(x)$ is a one-dimensional Wiener process. In the case of a k-dimensional parameter $x = ({}_jx, j = 1, \cdots, k)$, by a Brownian bridge $B(x)$ we mean

$$B(x) = W(x) - {}_1x\, {}_2x \cdots {}_kx W(1, 1, \cdots, 1)\,, \tag{4}$$

$0 \leqq {}_jx \leqq 1, j = 1, \cdots, k$. Here $W(x)$ is a k-dimensional Wiener process, that is, W is Gaussian with

$$\begin{aligned} &EW(x) \equiv 0 \\ &\operatorname{Cov}(W(x), W(x')) = \textstyle\prod_{j=1}^k \min({}_jx, {}_jx')\,, \end{aligned} \tag{5}$$

$0 \leqq {}_jx, {}_jx' \leqq 1, j = 1, \cdots, k$.

A result that fully corresponds to Theorem A in the k-dimensional case (in the sense that it is best) does not appear to be available yet. However, the following result due to Csörgö and Révész [3], [5] will be useful:

THEOREM B. *Let $X_1, X_2, \cdots, X_n, \cdots$ be independent random variables uniformly distributed over the unit k-cube $[0, 1]^k$. A sequence of Brownian bridges $B_n(x)$ can then be constructed so that*

$$\sup_{0 \leqq {}_jx \leqq 1; j=1,\cdots,k} |n^{\frac{1}{2}}(F_n(x) - \textstyle\prod_{j=1}^k {}_jx) - B_n(x)| = O(n^{-1/2(k+1)} (\log n)^{\frac{3}{2}}) \tag{6}$$

almost surely as $n \to \infty$.

The following result of Bickel and Rosenblatt [1] on the maximal deviation of a Gaussian process with k-variate parameter will be required in some of the derivation given later on:

THEOREM C. *Let $Y(x)$, $x = ({}_1x, \cdots, {}_kx)$, $-\infty < {}_jx < \infty, j = 1, \cdots, k$, be a Gaussian separable stationary process with mean zero and covariance function $r(x)$. Assume that*

$$r(x) = 1 - |x|^\alpha \textstyle\int_{S_k} |(x/|x|, \theta)|^\alpha \mu(d\theta) + o(|x|^\alpha)\,, \tag{7}$$

$0 < \alpha \leqq 2$, as $x \to 0$ with μ a finite measure on the unit k sphere S_k in R^k with at least two distinct rays in its spectrum so that the integral form on the right side of (7) is nonsingular. Further let $r \in L^2$. Set

$$\begin{aligned} B(t) = (2k \log t)^{\frac{1}{2}} + \Big\{\frac{1}{2}\Big(\frac{2k}{\alpha} - 1\Big) \log\log t \\ + \log((2\pi)^{-\frac{1}{2}} H_\alpha (2k)^{[(2k/\alpha)-1]/2})\Big\} (2k \log t)^{-\frac{1}{2}}\,, \end{aligned} \tag{8}$$

where

$$H_\alpha = \lim_{T\to\infty} T^{-k} \textstyle\int_0^\infty e^s P[\sup_{0 \leqq {}_1x, \cdots, {}_kx \leqq T} Z({}_1x, \cdots, {}_kx) > s]\, ds \tag{9}$$

with Z a Gaussian process having

$$E(Z(x)) = -|x|^\alpha \textstyle\int_{S_k} |(x/|x|, \theta)|^\alpha \mu(d\theta) = g(x)\,, \tag{10}$$

and

$$\operatorname{Cov}(Z(x), Z(x')) = g(x) + g(x') - g(x - x')\,. \tag{11}$$

Then if the maximal absolute value M_T *is*

(12) $$M_T = \max\{|Y(x)|, 0 \leqq {}_jx \leqq T, j = 1, \cdots, k\},$$

it follows that

(13) $$P[\{2k \log T\}^{\frac{1}{2}}(M_T - B(T)) < x] \rightarrow \exp\{-2\exp(-x)\},$$

as $T \rightarrow \infty$.

In the following two cases, H_α can be evaluated. If

(14) $$r(x) = 1 - x\Sigma x' + o(|x|^2)$$

with Σ a $k \times k$ nonsingular matrix, then

(15) $$H_2 = \pi^{-k/2}\{\det(\Sigma)\}^{\frac{1}{2}}.$$

If

(16) $$r(x) = 1 - C(|{}_1x| + \cdots + |{}_kx|) + o(\textstyle\sum_{j=1}^k |{}_jx|),$$

then

$$H_1 = C^k.$$

Theorem A implies the following theorem which is an improvement of Theorem 3.1 of [2]:

THEOREM 1. *Let* $k = 1$ *and assume the weight function* w *satisfies conditions* A1—A3 *of* [2] *with*

(17) $$b(n) = n^{-\delta}, \quad 0 < \delta < 1.$$

Then

(18) $$P\left[(2\delta \log n)^{\frac{1}{2}}\left(\frac{\tilde{M}_n}{(\lambda(w))^{\frac{1}{2}}} - d_n\right) < x\right] \rightarrow e^{-2e^{-x}},$$

with

(19) $$\lambda(w) = \int w^2(t)\,dt,$$

and

(20) $$d_n = (2\delta \log n)^{\frac{1}{2}} + \frac{1}{(2\delta \log n)^{\frac{1}{2}}}\left\{\log \frac{K_1(w)}{\pi^{\frac{1}{2}}} + \frac{1}{2}[\log \delta + \log\log n]\right\}$$

where

(21) $$K_1(w) = \frac{w^2(A) + w^2(-A)}{2\lambda(w)},$$

if $K_1(w) > 0$ *and otherwise*

(22) $$d_n = (2\delta \log n)^{\frac{1}{2}} + \frac{1}{(2\delta \log n)^{\frac{1}{2}}} \log \frac{1}{\pi}\left(\frac{K_2(w)}{2}\right)^{\frac{1}{2}},$$

where

(23) $$K_2(w) = \tfrac{1}{2}[\textstyle\int (w'(t))^2\,dt]/\lambda(w).$$

We make the following assumptions in the k-dimensional case which are referred to as assumptions B_1 and B_2.

B_1. The probability density f is positive and continuously differentiable up to kth order.

B_2. The weight function has bounded support and is continuously differentiable up to kth order.

The proof of a k-dimensional analogue of Theorem 1 will proceed via a series of intermediate approximations as in Theorem 3.1 of [2] though some variations will be required. Also, one result will not be as broad since Theorem B which is used does not give estimates as good as those supplied by Theorem A (or a corresponding result) in one dimension.

The process

$$Y_n(x) = n^{\frac{1}{2}} b(n)^{k/2} f(x)^{-\frac{1}{2}} (f_n(x) - Ef_n(x)) \tag{24}$$
$$= b(n)^{-k/2} f^{-\frac{1}{2}}(x) \int w\left(\frac{x-s}{b(n)}\right) dZ_n(s) ,$$

where

$$Z_n(s) = n^{\frac{1}{2}}(F_n(s) - F(s)) . \tag{25}$$

If the random variables ${}_1X, {}_2X, \cdots, {}_kX$ have joint distribution $F({}_1x, {}_2x, \cdots, {}_kx)$, let $F_1({}_1x)$ be the marginal distribution of ${}_1X$, $F_2({}_2x \mid {}_1x)$ the marginal conditional distribution of ${}_2X$ given ${}_1X, \cdots$, and $F_{k-1}({}_kx \mid {}_{k-1}x, \cdots, {}_1x)$ the conditional distribution of ${}_kX$ given ${}_{k-1}X, {}_{k-2}X, \cdots, {}_1X$. Let

$$\begin{aligned} {}_1X' &= F_1({}_1X) \\ {}_2X' &= F_2({}_2X \mid {}_1X) \\ &\cdots \\ {}_kX' &= F_{k-1}({}_kX \mid {}_{k-1}X, \cdots, {}_1X) , \end{aligned} \tag{26}$$

or

$$X' = ({}_1X', \cdots, {}_kX') = M({}_1X, {}_2X, \cdots, {}_kX) = MX \tag{27}$$

with M the transformation given by (26). If ${}_1X, \cdots, {}_kX$ have joint distribution F then ${}_1X', \cdots, {}_kX'$ are uniformly distributed on $[0, 1]^k$ given condition B_1 (see [6]). Using assumption B_2

$$\int w\left(\frac{x-s}{b(n)}\right) dZ_n(s) = \int w\left(\frac{x - M^{-1}s'}{b(n)}\right) dZ_n(M^{-1}s') \tag{28}$$
$$= \int \left(\frac{\partial}{\partial s_1'} \cdots \frac{\partial}{\partial s_k'}\right) \left\{w\left(\frac{x - M^{-1}s'}{b(n)}\right)\right\} Z_n(M^{-1}s')\, ds'$$

on integrating by parts. Theorem B implies that one can replace $Z_n(M^{-1}s')$ by a Brownian bridge $B(s')$ with a sufficiently small error. Thus we have

$$||Y_n(\cdot) - {}_0Y_n(\cdot)|| = O_p(b(n)^{-k/2} n^{-1/2(k+1)} (\log n)^{\frac{3}{2}}) \tag{29}$$

where $||\cdot||$ is the sup norm and

$$(30)\qquad {}_0Y_n(x) = b(n)^{-k/2}f^{-\frac{1}{2}}(x) \int \left(\frac{\partial}{\partial s_1'} \cdots \frac{\partial}{\partial s_k'}\right)\left\{w\left(\frac{x - M^{-1}s'}{b(n)}\right)\right\} B(s')\, ds'$$
$$= b(n)^{-k/2}f^{-\frac{1}{2}}(x) \int w\left(\frac{x-s}{b(n)}\right) dB(Ms) .$$

Further

$$(31)\qquad ||{}_0Y_n(\cdot) - {}_1Y_n(\cdot)|| = O_p(b(n)^{k/2}) ,$$

where

$$(32)\qquad {}_1Y_n(x) = b(n)^{-k/2}f^{-\frac{1}{2}}(x) \int w\left(\frac{x-s}{b(n)}\right) dW(Ms) ,$$

and W is k-dimensional Wiener process. This follows since f is bounded away from zero in $[0, 1]^k$. The process ${}_1Y_n(x)$ has the same probability structure as

$$(33)\qquad {}_2Y_n(x) = b(n)^{-k/2}f^{-\frac{1}{2}}(x) \int w\left(\frac{x-s}{b(n)}\right) f^{\frac{1}{2}}(s)\, dW(s) .$$

Let

$$(34)\qquad {}_3Y_n(x) = b(n)^{-k/2} \int w\left(\frac{x-s}{b(n)}\right) dW(s) .$$

Then since

$$(35)\qquad \begin{aligned} &{}_2Y_n(x) - {}_3Y_n(x) \\ &= b(n)^{-k/2} \int w\left(\frac{x-s}{b(n)}\right)\left\{\left(\frac{f(s)}{f(x)}\right)^{\frac{1}{2}} - 1\right\} dW(s) \\ &= b(n)^{-k/2} \int \left(\frac{\partial}{\partial s_1} \cdots \frac{\partial}{\partial s_k}\right)\left[w\left(\frac{x-s}{b(n)}\right)\left\{\left(\frac{f(s)}{f(x)}\right)^{\frac{1}{2}} - 1\right\}\right] W(s)\, ds , \end{aligned}$$

it follows that

$$(36)\qquad \sup_{0 \le {}_jx \le 1;\, j=1,\cdots,k} |{}_2Y_n(x) - {}_3Y_n(x)| = O_p(b(n)^{\frac{1}{2}}) .$$

The process ${}_3Y_n(b(n)x)$ has the same probability structure as the stationary Gaussian process

$$(37)\qquad U(x) = \int w(x-s)\, dW(s)$$

since it is Gaussian and has the same mean and covariance function as $U(x)$. The normalized process

$$(38)\qquad {}_1U(x) = U(x)/\{\lambda(w)\}^{\frac{1}{2}}$$

with $\lambda(w)$ given by (44) has covariance function

$$(39)\qquad \begin{aligned} r(x) &= \int w(x+y)w(y)\, dy/\lambda(w) \\ &= 1 - x\Sigma x' + o(|x|^2) \end{aligned}$$

as $x \to 0$ with

$$(40)\qquad \Sigma = \frac{1}{2}\left(\int \frac{\partial w}{\partial u_i}\frac{\partial w}{\partial u_j}\, du/\lambda(w)\right) .$$

If $b(n) = n^{-\delta}$ with $0 < \delta < 1/k(k+1)$, the analysis of the asymptotic distribution of (2) as $n \to \infty$ can be reduced to that of

$$\max_{0 < {}_jx < b(n)^{-1}; j=1,\cdots,k} |U_1(x)| \tag{41}$$

as $n \to \infty$. However, Theorem C can now be used to obtain the following theorem.

THEOREM 2. *Let f and w satisfy assumptions* B_1 *and* B_2 *with*

$$b(n) = n^{-\delta}, \quad 0 < \delta < \frac{1}{k(k+1)} . \tag{42}$$

Then

$$P\left[\{2k \log n^{\delta}\}^{\frac{1}{2}} \left(\frac{\tilde{M}_n}{(\lambda(w))^{\frac{1}{2}}} - d_n\right) < x\right] \to e^{-2e^{-x}} \tag{43}$$

as $n \to \infty$ where

$$\lambda(w) = \int w(u)^2 \, du , \tag{44}$$

and

$$d_n = \{2k \log n^{\delta}\}^{\frac{1}{2}} + \{\tfrac{1}{2}(k-1) \log \log n^{\delta} + \log ((2\pi)^{-\frac{1}{2}} H_2 (2k)^{(k-1)/2})\}(2k \log n^{\delta})^{-\frac{1}{2}} \tag{45}$$

with

$$H_2 = \pi^{-k/2} \{\det \Sigma\}^{\frac{1}{2}} , \tag{46}$$

(Σ *is given by* (40)).

With the error term given by Theorem B rather stringent conditions on f and w are required to replace $Ef_n(x)$ by $f(x)$ in Theorem 2. One certainly believes better conditions are possible. However, we shall state a corollary making use of stringent conditions. The additional condition B_3 is first required.

B_3 w satisfies the moment conditions

$$\int u_i^s w(u) \, du = 0 \tag{47}$$

for $s = 1, \cdots, l-1$ and $i = 1, \cdots, k$.

COROLLARY. *Let f and w satisfy the conditions of Theorem 2. Further assume w satisfies condition* B_3 *and that f is continuously differentiable up to order l with $l > 1/2\delta + k/2$. Then the conclusion of theorem holds with $\tilde{M}_n$ replaced by*

$$\tilde{M}_n = \max_{0 \leqq {}_jx \leqq 1; j=1,\cdots,k} n^{\frac{1}{2}} b(n)^{k/2} f(x)^{-\frac{1}{2}} |f_n(x) - f(x)| . \tag{48}$$

The corollary follows immediately on noting that

$$Ef_n(x) - f(x) = O(b(n)^l) . \tag{49}$$

We just remark in passing that a corresponding result for a weight function

$$\begin{aligned} w(x) &= 1 \qquad \text{if } \ |{}_jx| \leqq \tfrac{1}{2} \quad j = 1, \cdots, k \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

could be obtained by making use of Theorem C with condition (16).

REFERENCES

[1] BICKEL, P. J. and ROSENBLATT, M. (1973). Two-dimensional random fields. *In Multivariate Analysis* III, Krishnaiah, ed. 3–15.

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

[3] CSÖRGO, M. and RÉVÉSZ, P. (1975). A new method to prove Strassen type laws of invariance principle II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31** 261–269.

[4] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's and the sample df I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131.

[5] RÉVÉSZ, P. (1976). On strong approximation of the multidimensional empirical process. To appear.

[6] ROSENLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math Statist.* **23** 470.

DEPARTMENT OF MATHEMATICS
P. O. BOX 109
UNIVERSITY OF CALIFORNIA
LA JOLLA, CALIFORNIA 92037