Z. Wahrscheinlichkeitstheorie verw. Geb. 7, 327—335 (1967)

Estimation of a Density Function at a Point

L. WEISS* and J. WOLFOWITZ**

Received October 10, 1966

1. Introduction

Let $X_1, \ldots, X_n$ be one-dimensional, independent, identically distributed chance variables, to be observed by the statistician, with unknown density function $f(\cdot)$. That is, $f(\cdot)$ belongs to some class of densities containing more than one element. Let A be any point; we shall find an estimator (more properly, a sequence of estimators) $\{\varphi_n(X_1, \ldots, X_n),\ n = 1, 2, \ldots\}$ of $f(A)$ under varying assumptions on $f(\cdot)$ in the neighborhood of A. That is, for each n $\varphi_n(X_1, \ldots, X_n)$ is a Borel-measurable function of its arguments and, we hope, has a high probability of being close to $f(A)$.

Let $\varepsilon_n > 0$ be a function of n to be discussed shortly. Let $V(\varepsilon_n)$ be the class of estimators of $f(A)$ which, for $n = 1, 2, \ldots$, are functions only of those X's which lie in the interval $(A - \varepsilon_n, A + \varepsilon_n)$. Up to the end of the argument in (2.10) we shall make the following two assumptions on the admissible class of estimators:

Assumption I. All the estimators to be considered are in the class $V(\varepsilon_n)$.

Assumption II. The function ε_n is $n^{-\alpha}$, $\alpha > 0$.

These assumptions are a limitation on the class of estimators considered and should certainly be removed or substantially weakened. Nevertheless, the estimators we shall obtain will be better than estimators hitherto given in the literature. Subject to our assumptions we will discuss the optimal value of α. After (2.11) we will violate our own Assumption I by making use of observations outside of the interval $(A - \varepsilon_n, A + \varepsilon_n)$ in order to cope with the problem of estimating $K(\varepsilon_n)$; see, also Remark 1 of Section 2 below.

In what follows we shall consider three cases (problems), i.e., three different sets of assumptions about the totality of possible f. Other cases, involving higher derivatives, can be treated similarly; the general method will be apparent. In each case we will assume that the statistician knows only that f is a member of a certain class of densities, respectively W_1, W_2, and W_3. We now proceed to describe these classes.

Each class will consist of all densities which satisfy certain boundedness conditions near A. We begin by describing the class W_2. Any density g in W_2 satisfies

---

* Research supported by National Science Foundation Grant No. GP 3783.

** Fellow of the John Simon Guggenheim Memorial Foundation. Research supported by the U. S. Air Force under Contract AF 18 (600) — 685 with Cornell University.

both the following conditions:

(1.1) $$a_1 \leqq g(A) \leqq a_1'$$

(1.2) in the interval $I = (A - h, A + h)$

the second derivative $g''(\cdot)$ exists, $|g''(A)| < a_2$, and, for all y in I,

$$g(y) = g(A) + (y - A)g'(A) + \frac{(y - A)^2}{2} g''(A) + \bar{g}(y)|y - A|^{2+a},$$

where $0 < a < 1$ and $|\bar{g}(y)| \leqq a_4$.

In the above h, a, a_1, a_1', a_2, and a_4 are positive constants which determine a class W_2; different constants determine a different class. However, *it is not necessary for the statistician to know these constants*, and the estimators to be obtained will not depend on them. It is sufficient for the statistician to know that f belongs to *some* class W_2. The same remarks apply to the classes W_1 and W_3 to be described now. As before, $h, a, a_1, a_1', a_2, a_3$ and a_4 are some positive constants.

The class W_1 is to consist of all densities which satisfy the following conditions: Any density g satisfies (1.1) and

(1.3) in the interval $I = (A - h, A + h)$

the first derivative $g'(\cdot)$ exists, and, for all y in I,

$$g(y) = g(A) + (y - A)g'(A) + \bar{g}(y)|y - A|^{1+a}$$

where $|\bar{g}(y)| \leqq a_4$ and $0 < a < 1$.

The class W_3 is to consist of all densities which satisfy the following conditions: Any density g satisfies (1.1) and

(1.4) in the interval $I = (A - h, A + h)$

the third derivative $g'''(\cdot)$ exists, $|g''(A)| < a_2$, $|g'''(A)| < a_3$, and, for all y in I,

$$g(y) = g(A) + (y - A)g'(A) + \frac{(y - A)^2}{2} g''(A) + \frac{(y - A)^3}{6} g'''(A) + \bar{g}(y)|y - A|^{3+a},$$

where $|\bar{g}(y)| \leqq a_4$ and $0 < a < 1$.

We shall find it convenient to employ the notation now to be described. The statement $\psi = O(n^s)$ is to mean that $|\psi n^{-s}|$ is bounded above uniformly in (positive integral) n and all g in whatever W_i is relevant. The statement $\psi = o(n^s)$ is to mean that $\psi n^{-s} \to 0$ as $n \to \infty$ uniformly for all g in the relevant W. The statement $\psi = \Omega(n^s)$ is to mean that $|\psi n^{-s}|$ is bounded above and below by positive numbers, uniformly in n and g in W. Finally, O_p, o_p, and Ω_p are to mean that O, o, and Ω, respectively, hold with a probability which can be chosen arbitrarily close to one.

The symbol $P\{\ \}$ will always mean the probability of the relation in braces, when f is the common density of the X's. Of course, f may be any density in the class under consideration. Consequently, an inequality like $P\{R_1\} > P\{R_2\}$ will mean that the probability of R_1 is greater than that of R_2 no matter what f (in the appropriate class) may be.

2. The Case W_2

In the interval $I = (A - h, A + h)$ write $f(x) = f(A)[1 + k(x - A)]$ and

$$K(\varepsilon_n) = \int_{-\varepsilon_n}^{\varepsilon_n} k(y)\,dy\,, \quad \text{for } n \text{ such that} \quad n^{-\alpha} < h\,.$$

Suppose first that $K(\varepsilon_n)$ is known to the statistician. Let $Y_1, \ldots, Y_N$ be those among $X_1, \ldots, X_n$ which lie in the interval $(A - \varepsilon_n, A + \varepsilon_n)$. The joint probability function of N at m and probability density function of $Y_1, \ldots, Y_m$ at $y_1, \ldots, y_m$ is

$$\begin{aligned} &\frac{n!}{m!(n-m)!}[f(A)(2\varepsilon_n + K(\varepsilon_n))]^m \times \\ &\times [1 - f(A)(2\varepsilon_n + K(\varepsilon_n))]^{n-m} \prod_{i=1}^{m} \frac{f(A)(1 + k(y_i - A))}{f(A)(2\varepsilon_n + K(\varepsilon_n))}. \end{aligned} \tag{2.1}$$

From this we obtain the maximum likelihood estimator $\hat{f}_n$ of $f(A)$ (the value of $f(A)$ maximizing (2.1)) to be

$$\hat{f}_n = \frac{N}{n(2\varepsilon_n + K(\varepsilon_n))}. \tag{2.2}$$

Obviously $E\hat{f}_n = f(A)$ and

$$\sigma^2(\hat{f}_n) = E(\hat{f}_n - f(A))^2 = \frac{f(A)[1 - (2\varepsilon_n + K(\varepsilon_n))f(A)]}{n(2\varepsilon_n + K(\varepsilon_n))} = \Omega(n^{\alpha-1}). \tag{2.3}$$

Assume temporarily that $\alpha < 1/2$; it will turn out that anyhow α should be $\leqq 1/5$. One verifies easily that the distribution of

$$(\hat{f}_n - f(A))[\sigma(\hat{f}_n)]^{-1} \tag{2.4}$$

approaches the normal distribution with mean zero and variance one. By ESSEEN's theorem (e.g., [1]) for third moments it follows that the approach to this normal distribution is uniform in the argument of the limiting distribution and in the densities of W_2. The normalizing factor $[\sigma(\hat{f}_n)]^{-1}$ is $\Omega(n^{(1-\alpha)/2})$, and the chance variable $N = \Omega_p(n^{1-\alpha})$. It follows from Theorem 3.1 of [2] (see also [3]) that $\hat{f}_n$ is asymptotically efficient in the sense that it satisfies (3.8) of [2] for all competing estimators which satisfy (3.7) of [2] and Assumptions I and II of the present paper. Thus, if T_n is any such competing estimator we have, for any $r > 0$,

$$\begin{aligned} &\lim P\{-r n^{(\alpha-1)/2} < \hat{f}_n - f(A) < r n^{(\alpha-1)/2}\} \geqq \\ &\geqq \limsup P\{-r n^{(\alpha-1)/2} < T_n - f(A) < r n^{(\alpha-1)/2}\}. \end{aligned} \tag{2.5}$$

We now intend to cope with the problem created by the fact that $K(\varepsilon_n)$ is unknown. Since f is in W_2 we have that, for $-h < y < h$,

$$k(y) = k_1 y + k_2 y^2 + O(y^{2+a}) \tag{2.6}$$

and $k_2 = O(1)$. Hence

$$K(\varepsilon_n) = \left(\frac{2k_2}{3} + o(1)\right)\varepsilon_n^3. \tag{2.7}$$

Suppose we have an estimator $\hat{k}_2$ of k_2, and let

$$\hat{K}(\varepsilon_n) = \frac{2\hat{k}_2}{3}\varepsilon_n^3 \tag{2.8}$$

and

$$\hat{f}_n' = \frac{N}{n(2\varepsilon_n + \hat{K}(\varepsilon_n))}. \tag{2.9}$$

Define

$$D_n = K(\varepsilon_n) - \hat{K}(\varepsilon_n) = \left[\frac{2}{3}(k_2 - \hat{k}_2) + o(1)\right]\varepsilon_n^3. \tag{2.10}$$

Then we have

$$\hat{f}_n' - \hat{f}_n = \frac{N D_n}{n}[2\varepsilon_n + K(\varepsilon_n)]^{-1}[2\varepsilon_n + K(\varepsilon_n) - D_n]^{-1}. \tag{2.11}$$

To obtain $\hat{k}_2$ we proceed as follows: Let $J = (A - n^{-\beta}, A + n^{-\beta})$, $0 < \beta < 1/5$ to be determined later. Let $Z_1, \ldots, Z_{M(n)}$ be those of $X_1, \ldots, X_n$ which fall into the interval J.

Define

$$Q_n = \frac{\sum_1^{M(n)} |Z_i - A|}{M(n)}$$

and

$$\hat{k}_2 = 12\, n^{2\beta}(n^\beta Q_n - \tfrac{1}{2}). \tag{2.12}$$

The conditional density at the point $x = y + A$ of the interval J is

$$\frac{1 + k_1 y + k_2 y^2 + o(1) y^2}{2n^{-\beta} + \frac{2}{3}k_2 n^{-3\beta} + o(n^{-3\beta})}. \tag{2.13}$$

Hence

$$E|Z_i - A| = \frac{n^{-\beta}}{2}\left(1 + \frac{k_2}{6}n^{-2\beta} + o(n^{-2\beta})\right) \tag{2.14}$$

and

$$E(Z_i - A)^2 = \frac{n^{-2\beta}}{3}(1 + O(n^{-2\beta})) \tag{2.15}$$

Therefore

$$Q_n = \frac{n^{-\beta}}{2}\left(1 + \frac{k_2}{6}n^{-2\beta} + o(n^{-2\beta})\right) + \Omega_p(n^{-(1+\beta)/2}) \tag{2.16}$$

and

$$\hat{k}_2 = k_2 + \Omega_p(n^{-(1-5\beta)/2}) + o(1). \tag{2.17}$$

Then

$$D_n = \Omega_p(n^{-(1/2 - 5\beta/2 + 3\alpha)}) + o(n^{-3\alpha}) \tag{2.18}$$

and

$$\hat{f}_n' - \hat{f}_n = \Omega_p(n^{-(1/2 - 5\beta/2 + 2\alpha)}) + o_p(n^{-2\alpha}). \tag{2.19}$$

If we chose $\alpha = \frac{1}{5}$ and $\beta = \frac{1}{10}$ then

$$\hat{f}_n' - \hat{f}_n = O_p(n^{-13/20}) + o_p(n^{-2/5}) \tag{2.20}$$

while

$$\hat{f}_n = f(A) + \Omega_p(n^{-2/5}) \tag{2.21}$$

Thus $\hat{f}'_n$ is as efficient as $\hat{f}_n$ when $\alpha = 1/5$ and $\beta = 1/10$. (Actually when $\alpha = 1/5$, β can take any value in a suitable interval.) In this case, for any competing estimator T_n such as has been described earlier, we have, from (2.5),

$$\lim P\{-rn^{-2/5}<\hat{f}'_n-f(A)<rn^{-2/5}\}\geqq\limsup P\{-rn^{-2/5}< T_n-f(A)<rn^{-2/5}\}, \tag{2.22}$$

and the left member is positive by (2.20) and (2.21). It is remarkable that (2.22) is true, although the estimator T_n may be explicitly a function of $K(\varepsilon_n)$, while $\hat{f}'_n$, of course, is not.

We turn now to the problem of how to choose the α with which $\hat{f}'_n$ is to be computed. Suppose first that $\hat{f}'_n$ is computed with an $\alpha > 1/5$. Then, by (2.3) and (2.19), we have

$$\hat{f}'_n = f(A) + \Omega_p(n^{(\alpha-1)/2}),$$

so that, by (2.20) and (2.21), this choice of α is worse than that of $\alpha = 1/5$. Suppose now that $\alpha = 1/5 - d$, $0 < d < 1/5$. In order to perform a finer computation we replace the density (2.13) by

$$\frac{1 + k_1 y + k_2 y^2 + O(y^{2+a})}{2n^{-\beta} + \frac{2}{3}k_2 n^{-3\beta} + O(n^{-\beta(3+a)})}.$$

The last term of the right member of (2.17) is to be replaced by $O(n^{-a\beta})$, and that in (2.19) by

$$O_p(n^{-a\beta-2\alpha}) + O_p(n^{-(2+a)\alpha}) = O_p(n^{-a\beta+2d-2/5}) + O_p(n^{-2/5-a/5+2d+ad}). \tag{2.23}$$

We want to maintain, for any $r > 0$, the validity of

$$\begin{aligned}&\lim P\{-rn^{(\alpha-1)/2} < \hat{f}'_n - f(A) < rn^{(\alpha-1)/2}\}\\ &\qquad= \lim P\{-rn^{(\alpha-1)/2} < \hat{f}_n - f(A) < rn^{(\alpha-1)/2}\},\end{aligned} \tag{2.24}$$

because then, from (2.5),

$$\begin{aligned}&\lim P\{-rn^{(\alpha-1)/2} < \hat{f}'_n - f(A) < rn^{(\alpha-1)/2}\} \geqq\\ &\qquad\geqq \limsup P\{-rn^{(\alpha-1)/2} < T_n - f(A) < rn^{(\alpha-1)/2}\},\end{aligned} \tag{2.25}$$

and, by (2.3) and (2.24), the left member of (2.25) is positive. Here T_n is any competing estimator such as has been described earlier, and, in the definition of T_n, the statistician may even assume that $K(\varepsilon_n)$ is known to him.

In order for (2.24) to hold it follows from (2.3), (2.19), and (2.23) that we must have

$$d < \frac{2\beta a}{5}. \tag{2.26}$$

$$d < \frac{2a}{5(5+2a)}, \tag{2.27}$$

and

$$\beta < \alpha. \tag{2.28}$$

The inequality (2.28) can always be achieved by proper choice of β, and it is (2.26) and (2.27) which require attention. We consider two possibilities:

A) The realistic situation is that the statistician does not know a at all. In that case he will choose $d = 0 (\alpha = 1/5)$ in order to be certain that (2.26) and (2.27) hold. Then (2.22) obtains.

B) The statistician knows a and chooses d and β so as to satisfy (2.26)—(2.28). The result (2.25) is improved because α is smaller.

We have assumed that $0 < a < 1$. This seems entirely reasonable and natural. However, if $a > 1$ the reader will verify very easily that then the only change required in the above argument is that (2.26) be replaced by

$$d < \frac{2\beta}{5}.$$

Remark 1. Since $\beta < \alpha$ it follows that, to estimate $K(\varepsilon_n)$, we have employed observations outside of the interval $(A - n^{-\alpha}, A + n^{-\alpha})$ and violated our own Assumption I. It seems quite certain that we have not made full use of all the observations in $(A - n^{-\beta}, A + n^{-\beta})$. Indeed, β was not even uniquely determined. It seems clear to us that it must be possible to improve our method. On behalf of the latter it must be said that it works and, where comparison is possible, usually gives better results than hitherto given methods. (See, for example, Remarks 2, 3, and 4 below.) Most papers in the literature ignore the question of efficiency. Consistent estimators are easy to give.

Remark 2. The estimator most commonly given in the literature is

$$\bar{f}_n = \frac{N}{2n\varepsilon_n}.$$

When comparing $\hat{f}'_n$ with $\bar{f}_n$ we must bear in mind that $\bar{f}_n$ might show up to better advantage at different values of α from those best for $\hat{f}'_n$. To avoid confusion, we write $\bar{\alpha}$ instead of α when referring to $\bar{f}_n$. In the present remark we will always assume that f is in class W_2.

It is easily shown that $\bar{f}_n$ is asymptotically normal with mean

$$f(A)\left[1 + \frac{n^{-2\bar{\alpha}}}{2}\left\{\frac{2}{3}k_2 + o(1)\right\}\right]$$

and standard deviation

$$\sqrt{\frac{f(A)}{2}}\, n^{-(1-\bar{\alpha})/2}(1 + o(1)].$$

For a given positive r, denote

$$P[-rn^{-2/5} < \bar{f}_n - f(A) < rn^{-2/5}]$$

by $p_n(f, \bar{\alpha}, r)$.

We will compare the estimator $\bar{f}_n$ with the estimator $\hat{f}'_n$ computed using $\alpha = 1/5$ and $\beta = 1/10$. From (2.20), we know that $\hat{f}'_n$ is asymptotically normal, with mean $f(A)$ and standard deviation $\sqrt{\frac{f(A)}{2}}\, n^{-2/5} \times [1 + o(1)]$. Denote

$$P[-rn^{-2/5} < \hat{f}'_n - f(A) < rn^{-2/5}]$$

by $q_n(f, r)$.

The following facts are now easily shown:

$$\lim_{n\to\infty} q_n(f, r) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\sqrt{\frac{2}{f(A)}}\,r}^{\sqrt{\frac{2}{f(A)}}\,r} \exp\left\{-\frac{t^2}{2}\right\} dt = L(f, r), \quad \text{say}.$$

If $\bar{\alpha} > 1/5$, $\lim_{n\to\infty} p_n(f, \bar{\alpha}, r) = 0$. If $\bar{\alpha} = 1/5$, $\lim_{n\to\infty} p_n(f, \bar{\alpha}, r)$ exists and is less than $L(f, r)$ unless $k_2 = 0$, when it is equal to $L(f, r)$. If

$$\bar{\alpha} < 1/5 \quad \text{and} \quad k_2 \neq 0, \quad \lim_{n\to\infty} p_n(f, \bar{\alpha}, r) = 0.$$

Since k_2 is not known, and the estimator $\bar{f}_n$ makes no attempt to estimate it, $\bar{f}_n$ in most situations is inferior to $\hat{f}'_n$.

3. The Case W_1

We will try to use the argument and notation of Section 2 as much as possible and indicate chiefly the necessary changes. This case is very simple. We define $\hat{f}'_n$ as $\bar{f}_n$, and $\bar{f}_n$ and $\hat{f}_n$ are defined exactly as in Section 2. The changes in this case are in the estimation of $K(\varepsilon_n)$ and in the choice of α. As before, (2.3) and (2.5) hold. Now we have

$$k(y) = k_1 y + O(y^{1+a}). \tag{3.1}$$

Hence

$$K(\varepsilon_n) = O(\varepsilon_n^{2+a}) \tag{3.2}$$

and

$$\bar{f}_n - \hat{f}_n = O(\varepsilon_n^{1+a}). \tag{3.3}$$

Since (2.3) holds we need that

$$-(1 + a)\alpha < \frac{\alpha - 1}{2}. \tag{3.4}$$

If the statistician does not know a then he should choose $\alpha = 1/3$. If he knows a then he should choose

$$\alpha > \frac{1}{3 + 2a}. \tag{3.5}$$

The smaller α the smaller is $\sigma(\bar{f}_n)$.

4. The Case W_3

We define $\hat{f}_n$, $\hat{K}(\varepsilon_n)$, $\hat{k}_2$, and $\hat{f}'_n$ as in Section 2. Write

$$k(y) = k_1 y + k_2 y^2 + k_3 y^3 + O(y^{3+a}).$$

Then

$$K(\varepsilon_n) = \frac{2k_2}{3} \varepsilon_n^3 + O(\varepsilon_n^{4+a}). \tag{4.1}$$

The last term in the parentheses in the right member of (2.14) can now be replaced by $O(n^{-\beta(3+a)})$. In place of (2.17) we now have

$$\hat{k}_2 = k_2 + \Omega_p(n^{-(1-5\beta)/2}) + O(n^{-\beta(1+a)}). \tag{4.2}$$

In place of (2.18) we now have

$$\begin{aligned} D_n = & \Omega_p(n^{-1/2-5\beta/2+3\alpha}) \\ & + O(n^{-\beta(1+a)-3\alpha}) \\ & + O(n^{-(4+a)\alpha}). \end{aligned} \tag{4.3}$$

In place of (2.19) we now have

$$\begin{aligned} \hat{f}'_n - \hat{f}_n = & \Omega_p(n^{-(1/2-5\beta/2+2\alpha)}) \\ & + O_p(n^{-\beta(1+a)-2\alpha}) \\ & + O_p(n^{-(3+a)\alpha}). \end{aligned} \tag{4.4}$$

Our choice of α is the smallest value for which

$$\hat{f}'_n - \hat{f}_n = o_p(n^{(\alpha-1)/2}). \tag{4.5}$$

The latter is assured if all of the following hold:

$$\alpha > \beta, \tag{4.6}$$

$$5\alpha + 2\beta > 1 - 2\beta a, \tag{4.7}$$

$$7\alpha > 1 - 2a\alpha. \tag{4.8}$$

If a is known then one chooses β and the conveniently smallest α so that (4.6) —(4.8) are satisfied. In the more realistic situation where a is not known a satisfactory choice for α is a number a little larger than $1/7$, and β such that $\alpha > \beta > 1/7$. Writing $\alpha = 1/7 + \delta$, $\delta > 0$, we then have

$$\hat{f}'_n = f(A) + \Omega_p(n^{-3/7+\delta/2}). \tag{4.9}$$

In any case we can always choose $\delta < 2/35$ ($\alpha < 1/5$), improving the result for $\alpha = 1/5$ obtained in Section 2.

Remark 3. Again we compare the behavior of $\hat{f}'_n$ with that of the popular $\bar{f}_n$ in the case W_3. Any class W_3 is a subclass of suitable W_2 classes. We have already seen for what densities f in W_2 the estimator $\hat{f}'_n$ (computed with $\alpha = 1/5$), is more efficient than $\bar{f}_n$, and that the latter is the case in most situations. The same conclusions therefore apply in the case W_3.

Remark 4. In [4] the estimator $\bar{f}_n$ is discussed under conditions which are not entirely explicitly given, but seem to be very close to those of case W_3. The value $\alpha = 1/5$ ($\delta = 2/35$) seems to be recommended.

In [5] a generalization of $\bar{f}_n$ is discussed, but there is no discussion of optimality.

In [6] the problem of finding an estimator $f_n(x)$ which minimizes the expected value of $\int_{-\infty}^{\infty} [f_n(x) - f(x)]^2 dx$ was discussed.

5. Concluding Remarks

In all of the above we took A to be the center of the interval of length $2\varepsilon_n$; this is of course not necessary. Suppose the density does not exist at A, but there are right and left densities (right and left derivatives of the distribution function). The methods of the present paper are applicable to the estimation of these densities; one takes A to be the appropriate end-point of the interval whose observations are employed. The same is true if appropriate derivatives of the density exist only from the left or only from the right.

Application of the present methods will yield an estimator of the value of a multi-dimensional density at a point.

References

1. Esseen, C. G.: Fourier analysis of distribution functions. Acta math. **77**, 1—125 (1944).
2. Weiss, L. and J. Wolfowitz: Generalized maximum likelihood estimators. Teor. Verojatn. Primen, XI, 68—93 (1966).
3. Wolfowitz, J.: Asymptotic efficiency of the maximum likelihood estimator. Teor. Verojatn. Primen X, 267—281 (1965).
4. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. Ann. math. Statistics **27**, 832—837 (1956).
5. Parzen, E.: On estimation of a probability density function and a mode. Ann. math. Statistics **33**, 1065—1076 (1962).
6. Watson, G. S., and M. R. Leadbetter: On the estimation of the probability density I. Ann. math. Statistics **34**, 480—491 (1963).

Cornell University
Ithaca, New York, U. S. A.