JOURNAL OF COMPUTATIONAL PHYSICS 74, 409–421 (1988)

The Kernel Method for Unfolding Sphere Size Distributions

PETER HALL

Department of Statistics, Faculty of Economics, Australian National University, Canberra, ACT 2601, Australia

AND

RICHARD L. SMITH

Department of Mathematics, University of Surrey, Guildford GU2 5XH, England

Received June 12, 1986; revised October 20, 1986

Suppose spherical particles are embedded in an opaque matrix. Sphere radii are independent random variables with unknown density f. We wish to estimate f from the circular profiles of the spheres intersecting a plane section through the material. A recent proposal by C. C. Taylor (*J. Microscopy* **132**, 57 (1983)) is to estimate f by stereologically unfolding a kernel estimator derived from the observed profiles. We give some theoretical results for the bias and variance of this estimator and draw some sharp contrasts with ordinary density estimation. The unfolding causes the estimator to converge at a surprisingly slow rate, but it is also shown that this slow convergence is intrinsic to the problem and is not a deficiency of Taylor's method. The arguments are extended to cover also some other cases of stereological unfolding where a theoretical closed-form solution exists. © 1988 Academic Press, Inc.

1. INTRODUCTION

Suppose that spherical particles are embedded in an opaque matrix. The radii are independent and identically distributed random variables, with unknown probability density f. We wish to estimate f from the observed circular profiles of the particles in a cross section through the material. This so-called "unfolding" problem is a very well-known problem of stereology, see, e.g., Coleman [3] or Ripley [9, Chap. 9] for reviews.

Recently Taylor [12] proposed a new solution to this problem based on the kernel method for nonparametric density estimation. He presented examples of the practical implementation of his method, but he did not discuss such theoretical properties as the bias and variance of the estimator. Our purpose here is to give an outline of how such results may be derived and to draw attention to a surprising way in which they differ from the corresponding results for ordinary nonparametric density estimation. The practical consequences of these results are that a smoothing parameter called the "window width" or "band width" must be chosen considerably

0021-9991/88 $3.00

larger than in ordinary nonparametric density estimation, and there is a corresponding loss of accuracy in the estimator itself. Moreover, this loss of accuracy is not a deficiency of Taylor's particular procedure, but in a certain well-defined sense can be shown to be inherent to the problem of stereological unfolding.

Let f denote the density of sphere radius, with $m = \int xf(x)\,dx$ the mean sphere radius, and let g denote the corresponding density of random cross-sectional radius. Then g is derived from f by the Abel integral equation

$$g(y) = m^{-1} y \int_y^{\infty} (x^2 - y^2)^{-1/2} f(x)\,dx, \tag{1.1}$$

with inverse

$$f(x) = -2\pi^{-1} m \frac{d}{dx} \left\{ \int_x^{\infty} (y^2 - x^2)^{-1/2} g(y)\,dy \right\}. \tag{1.2}$$

Equations (1.1) and (1.2) are standard equations for this problem; note that the assumption of independent sphere centres is not strictly consistent with the notion of nonoverlapping spheres, but these equations are valid if overlapping is permitted or approximately valid if there is no overlapping but the sphere density is low. Taylor's [12] method was first to estimate g by the kernel method

$$\hat{g}(y) = (nh)^{-1} \sum_{i=1}^{n} K(h^{-1}(y - Y_i)) \tag{1.3}$$

and then to apply (1.2) to obtain an estimate $\hat{f}$ of f. In (1.3), $Y_1, ..., Y_n$ are the observed radii of circular profiles, K is a kernel function (usually a nonnegative function satisfying $\int K(x)\,dx = 1$) and h is the window width. The choice of window width controls the smoothness of the resulting estimate, and is well known to be an important factor in nonparametric density estimation (see, e.g., Wegman [13], Fryer [5]). We shall show that this choice is also important here, but that both the optimal window width and the corresponding rate of convergence of the estimator are different from ordinary density estimation.

As well as Taylor's procedure, we shall consider a variant which is easier to analyse mathematically, and perhaps more natural in practice as well. Define $f_1(x) = (2x^{1/2})^{-1} f(x^{1/2})$, $g_1(y) = (2y^{1/2})^{-1} g(y^{1/2})$; these are the densities of squared radii for spheres and circles, respectively. Then (1.1) and (1.2) become

$$g_1(y) = (2m)^{-1} \int_y^{\infty} (x - y)^{-1/2} f_1(x)\,dx, \tag{1.4}$$

$$f_1(x) = -2\pi^{-1} m \frac{d}{dx} \left\{ \int_x^{\infty} (y - x)^{-1/2} g_1(y)\,dy \right\}. \tag{1.5}$$

This suggests estimating g_1 by a Kernel estimator

$$\hat{g}_1(y) = (nh)^{-1} \sum_{i=1}^{n} K(h^{-1}(y - Y_i^2)) \tag{1.6}$$

and then applying (1.5) to estimate f_1 by $\hat{f}_1$. This is more convenient mathematically because the integrals in (1.4) and (1.5) are convolution integrals. The practical motivation is that Y_i^2 is proportional to the observed cross-sectional area, which may be easier to measure than the radius.

For alternative approaches to the unfolding problem, see in particular Jakeman and Anderssen [7], Anderssen and Jakeman [1], Kanatani and Ishikawa [8]. Jakeman and Anderssen consider, not only the above classical unfolding problem, but also a number of related problems for which closed-form solutions exist. These problems originated in the work of Santaló [10]. Kanatani and Ishikawa [8] consider a number of computational approaches both for the classical problem described above and for an extension which is relevant when the cross section consists of thin slice, but they do not consider the kernel method which, in our opinion, is a fully viable alternative to the schemes they describe. Some of these related problems will be studied in Section 4. For details of the numerical procedure, we refer to Taylor [12]. It is possible to carry out the inversion, (1.2) or (1.5), analytically provided the kernel is chosen appropriately, but the computation of the kernel density itself may be speeded up by the use of the fast Fourier transform if the sample size is very large. Taylor recommends truncating the inversion at the smallest order statistic and then rescaling so that the integral of the density is 1. He also describes a variant of the kernel procedure in which the window width varies over the sample, but we shall not consider that here.

2. Bias and Variance of the Unfolded Kernel Estimate

For simplicity we shall first consider the estimator defined by (1.5) and (1.6), i.e.,

$$\hat{f}_1(x) = -2(n\pi h^2)^{-1} m \sum_{i=1}^{n} \int_x^{\infty} (y - x)^{-1/2} K'(h^{-1}(y - Y_i^2))\, dy. \tag{2.1}$$

At this stage we assume that m, the mean of f, is known; later we shall remove that assumption. The mean of $\hat{f}_1$ is

$$E\{\hat{f}_1(x)\} = -2(\pi h^2)^{-1} m \int_{-\infty}^{\infty} \int_x^{\infty} (y - x)^{-1/2} K'(h^{-1}(y - v))\, g_1(v)\, dy\, dv$$

and, after substituting from (1.4), becomes

$$E\{\hat{f}_1(x)\} = \int_{-\infty}^{\infty} K(z) f_1(x - zh)\, dz. \tag{2.2}$$

Although it is possible to derive (2.2) just by algebraic manipulation, it may help to give a quick "symbolic" proof. We may rewrite (1.4) and (1.5) as

$$g_1 = Tf_1, \qquad f_1 = T^{-1} g_1,$$

where T is an operator and T^{-1} is its inverse. Moreover, the expression

$$E\{\hat{g}_1(y)\} = h^{-1} \int_{-\infty}^{\infty} K(h^{-1}(y-v))\, g_1(v)\, dv = \int_{-\infty}^{\infty} K(z)\, g_1(y-zh)\, dz$$

may also be thought of as an operator acting on g_1; thus $E\{\hat{g}_1\} = Sg_1$, say. Combining these formulae,

$$E\{\hat{f}_1\} = E\{T^{-1}\hat{g}_1\} = T^{-1}STf_1. \tag{2.3}$$

Now the operators S, T, and T^{-1} are all *convolution* operators and therefore *commute*. In other words, we can interchange the order of S and T in (2.3) and so cancel T with T^{-1}. Equation (2.3) then reduces to $E\{\hat{f}_1\} = Sf_1$, which is (2.2).

Now suppose that the kernel function K satisfies

$$\int_{-\infty}^{\infty} x^j K(x)\, dx = \begin{cases} 1, & j=0, \\ 0, & j=1, ..., k-1, \\ d_k \neq 0, & j=k, \end{cases} \tag{2.4}$$

for some integer $k > 1$. In most practical cases, in particular if K is a nonnegative function, we will have $k = 2$, but there are some theoretical grounds [2] for preferring a kernel with $k > 2$. In this case, taking the first k terms of a Taylor expansion of f and applying (2.4) to (2.2), we have

$$E\{\hat{f}_1(x)\} - f_1(x) = \frac{(-h)^k\, d_k f_1^{(k)}(x)}{k!} + o(h^k). \tag{2.5}$$

So far, our results are identical with those for ordinary density estimation, in which we have a sample directly from the density f_1; see, e.g., Bartlett [2], Singh [11].

Now let us consider the variance of $\hat{f}_1$. If we define a new kernel

$$K_1(x) = \int_0^{\infty} y^{-1/2} K'(x+y)\, dy, \qquad -\infty < x < \infty,$$

then we may write

$$\hat{f}_1(x) = -2(n\pi h^{3/2})^{-1}\, m \sum_{i=1}^{n} K_1(h^{-1}(x - Y_i^2))$$

from which we deduce

$$n\,\mathrm{var}\{\hat{f}_1(x)\} = 4\pi^{-2}h^{-2}m^2 \int_{-\infty}^{\infty} K_1^2(v)\, g_1(x-vh)\, dv - \{E\hat{f}_1(x)\}^2. \tag{2.6}$$

The second term tends to $f_1^2(x)$ as $h \to 0$, and hence is negligible compared with the first. Define

$$C_K = 4\int_{-\infty}^{\infty} K_1^2(v)\, dv = \int_0^{\infty}\int_{-\infty}^{\infty} \left\{\frac{K(v+z)-K(v)}{z}\right\}^2 dv\, dz. \tag{2.7}$$

Substituting from (2.7) in (2.6) and letting $h \to 0$, we obtain

$$n\,\mathrm{var}\{\hat{f}_1(x)\} = \pi^{-2}h^{-2}m^2 C_K\, g_1(x) + o(h^{-2}) \tag{2.8}$$

as $h \to 0$. It is here that we see a substantial difference from ordinary density estimation. In ordinary density estimation, the variance turns out to be of the order of $n^{-1}h^{-1}$ whereas here it is of the order of $n^{-1}h^{-2}$.

We now consider the consequences of these results for mean squared error. From (2.5) and (2.8) we have

$$\begin{aligned} E\{\hat{f}_1(x) - f_1(x)\}^2 &= \mathrm{var}\{\hat{f}_1(x)\} + \{E\hat{f}_1(x) - f_1(x)\}^2 \\ &\approx \alpha_1 n^{-1}h^{-2} + \alpha_2 h^{2k}, \end{aligned} \tag{2.9}$$

where $\alpha_1 = \pi^{-2}m^2 C_K\, g_1(x)$ and $\alpha_2 = \{d_k f_1^{(k)}(x)/k!\}^2$. The right-hand side of (2.9) is minimised by $h = h_1^* = \beta_1 n^{-1/(2k+2)}$, where $\beta_1 = (\alpha_1/k\alpha_2)^{1/(2k+2)}$. The corresponding mean squared error is of $O(n^{-2k/(2k+2)})$. In ordinary density estimation, the asymptotically optimal window width is a constant multiple of $n^{-1/(2k+1)}$ and the corresponding mean squared error is of $O(n^{-k/(2k+1)})$ [11]. Thus the optimal window width and mean squared error are both a larger order of magnitude in the unfolding problem.

Similar results hold for Taylor's estimator $\hat{f}$, defined by combining (1.2) and (1.3). In this case we find that

$$E\{\hat{f}(x)\} - f(x) = \frac{(-h)^k\, d_k f^{(k)}(x)}{k!} + o(h^k), \tag{2.10}$$

$$n\,\mathrm{var}\{\hat{f}(x)\} \approx 2\pi^{-2}h^{-2}m^2 C_K x^{-1} g(x). \tag{2.11}$$

These are the same orders of magnitude as for $\hat{f}_1$, and thus lead to similar conclusions about the optimal window width and mean squared error.

The derivation of (2.10) and (2.11) is considerably more complicated than that of (2.5) and (2.8), and we omit all technical details. A technical report is available from the second author, giving precise statements and proofs of these formulae. In Appendix 1, we list the assumptions required on both the kernel and the density function to make our results rigorous.

Estimating m. In our analysis so far, we have assumed that $m = \int xf(x)\,dx$ is known. In practice it will be unknown, but the formula

$$E\{Y^{-1}\} = \int y^{-1} g(y)\,dy = \pi/(2m)$$

[9, Chap. 9] suggests the estimator

$$\hat{m} = (\pi/2)\,n\left(\sum Y_i^{-1}\right)^{-1}.$$

It may be shown that

$$\hat{m} - m = O_p(n^{-1/2+\delta})$$

for any $\delta > 0$. In other words, the rate of convergence of $\hat{m}$ to m is arbitrarily close to order $n^{-1/2}$, which is a faster rate of convergence than that of $\hat{f}_1$ or $\hat{f}$. Consequently, $\hat{m}$ may be substituted for m in the definition of $\hat{f}_1$ or $\hat{f}$, with no change in the asymptotic formulae that have been derived.

3. Optimality

The main conclusion of Section 2 was that, when kernel estimation is applied to the unfolding problem, the rate of convergence is slower than in ordinary density estimation. From the discussion so far, it is unclear whether this is a peculiarity of kernel estimation or is inherent in the problem. In this section we state a theorem which indicates that the latter is the case and that there is a sense in which kernel estimation achieves the optimal rate of convergence. Proof of the theorem is deferred to Appendix 2.

In establishing optimality it suffices to confine attention to estimates of f_1. Analogous results for f follow by the obvious transformation.

For given $k \geq 1$ and $B > 0$, let $C_k(B)$ denote the class of densities f_1 with support confined to $(0, \infty)$ such that f_1 and its first $k-1$ derivatives exist on $(0, \infty)$. and $f_1^{(k-1)}$ is absolutely continuous on $(0, \infty)$, with the essential supremum of $f_1^{(j)}$ bounded by B for $0 \leq j \leq k$. Let $\phi_n(Z_1, ..., Z_n)$ denote a nonparametric estimate of $f_1(x_0)$, for some fixed $x_0 > 0$, based on a random sample $Z_1, ..., Z_n$ from the distribution with density g_1.

THEOREM. *Suppose, for some sequence of positive constants a_n, $n > 1$, we have*

$$\liminf_{n \to \infty} \inf_{f_1 \in C_k(B)} P_{g_1}\{|\phi_n(Z_1, ..., Z_n) - f_1(x_0)| \leq a_n\} = 1. \tag{3.1}$$

Then

$$\liminf_{n \to \infty} n^{k/(2k+2)} a_n = \infty. \tag{3.2}$$

The interpretation of this result is as follows. The estimator ϕ_n is an arbitrary estimator of $f_1(x_0)$; thus it could be our kernel estimator but it could also be something quite different. The class of densities $C_k(B)$ is effectively the set of all densities whose first k derivatives are bounded by B. Equation (3.1) implies that the error in ϕ_n as an estimate of $f_1(x_0)$ is at most a_n, uniformly over the class $C_k(B)$. Then (3.2) implies a lower bound on the rate at which $a_n \to 0$. In fact, for the kernel estimate with window width chosen to minimise the mean squared error, it may be shown that (3.2) implies (3.1), and in this sense the kernel estimate is optimal in the class of all density estimates based on $Z_1, ..., Z_n$.

4. Extension to Some Related Stereological Problems

Jakeman and Anderssen [7] consider a number of other unfolding problems, originating in the work of Santaló [10], for which a closed-form analytic solution exists. There has also been extensive work on the so-called "thin section" problem, in which the sphere is intercepted not by a plane but by a section of small but positive thickness; see Coleman [3, Section 4.3], Kanatani and Ishikawa [8], Jakeman [6]. The extension of the kernel method to these problems is immediate: the folded density is estimated by the kernel method from the observed data and is then unfolded by means of the analytic inversion formula. In this section we show that results, in terms of the optimal window width and corresponding mean squared error, hold for these problems which are similar to those for the classical unfolding problem described in Section 2. The mathematical treatment in this section is to some extent heuristic.

Let us first consider the various alternative schemes listed in Table 1 of Jakeman and Anderssen [7]. The "spherical" case of their (a)(i) is the classical problem already described. The "approximately spherical" case leads to the inversion formula

$$g(x) = -\frac{4a_n x^{2(1-\mu)} \sin(\mu\pi)\, N_A}{vCN_A} \int_{a_n x^2}^{\infty} \frac{z'(a)}{(a - a_n^2 x)^{1-\mu}}\, da, \tag{4.1}$$

where $z(a)$ is the observed density of cross-sectional area and a_n, μ, v, N_A, and C are constants. Evaluating this for fixed x and ignoring the constants, we see that it is again of the form (1.5) but with the power $-\frac{1}{2}$ replaced by $-1+\mu$. Suppose that we estimate $z(\cdot)$ by a kernel estimator

$$\hat{Z}(a) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{a - A_i}{h}\right)$$

where $A_1, ..., A_n$ are observed cross-sectional areas and K is a kernel function satisfying (2.4), and then substitute into (4.1) to estimate $g(\cdot)$. The mean squared error is minimised by taking h proportional to $n^{-1/(2k+3-2\mu)}$, and then becomes of $O(n^{-2k/(2k+3-2\mu)})$.

The other problems listed in Jakeman and Anderssen's Table 1 are the types already studied. Their problems (a)(ii) and (b)(i) both involve inversion formulae of the same kind as (1.2), and hence are also covered by the results of our Sections 2 and 3, while in their problems (b)(ii) and (c) the inversion formula involves nothing more than differentiation of the estimated density, and hence is covered by the results of Singh [11].

We now consider the "thin section" problem. As stated by Jakeman [6], the integral equation relating the folded and unfolded densities, g and f, is

$$(2a+t)\,g(y)=2y\int_y^\infty (x^2-y^2)^{-1/2} f(x)\,dx + tf(x), \tag{4.2}$$

where t is the thickness of the section. Note that Coleman [3] considers a further generalisation in which there is a truncation point, or smallest observable radius, below which a sphere is not observed at all. The inversion formula for (4.2) is

$$f(x)=-(2/\pi)^{1/2}(2a+t)\,t^{-1}x\int_x^\infty u\left\{\frac{2\pi(y^2-x^2)^{1/2}}{t}\right\}\frac{d}{dy}\left\{\frac{g(y)}{y}\right\}dy \tag{4.3}$$

(Jakeman's Eq. (7)), where

$$u(x)=e^{x^2/2}\int_x^\infty e^{-t^2/2}\,dt, \qquad a=\int_0^\infty xf(x)\,dx. \tag{4.4}$$

Writing $f(x)=2xf_1(x^2)$, $g(y)=2yg_1(y^2)$ as before, (4.2) and (4.3) become

$$(2a+t)\,g_1(y)=\int_y^\infty (x-y)^{-1/2} f_1(x) + tf_1(y), \tag{4.5}$$

$$f_1(x)=-(2/\pi)^{1/2}(2a+t)\,t^{-1}\int_x^\infty u\left\{\frac{2\pi(y-x)^{1/2}}{t}\right\}g_1'(y)\,dy. \tag{4.6}$$

We again assume that g_1 is estimated by $\hat{g}_1$, given by (1.6), and that this substituted into (4.6) to obtain an estimate $\hat{f}_1$.

We now make heuristic calculations of the bias and variance of this estimator at a specific value of x. Since the operator taking f_1 into g_1 is still a convolution operator, the argument of Section 2 again shows that the bias in $\hat{f}_1$ is given by (2.5). For the variance, define

$$K_h(x)=-(2/\pi)^{1/2}(2a+t)\,t^{-1}\int_0^\infty u\left\{\frac{2\pi(hy)^{1/2}}{t}\right\}K'(y+x)\,dy \tag{4.7}$$

and note that

$$\hat{f}_1(x) = (nh)^{-1} \sum_{i=1}^{n} K_h\{h^{-1}(x - Y_i^2)\}.$$

By a similar argument to that in Section 2,

$$n \operatorname{var} \hat{f}_1(x) \sim g_1(x)\, h^{-1} \int_{-\infty}^{\infty} K_h^2(z)\, dz.$$

We could now construct an asymptotic theory as $n \to \infty$, $h \to 0$ for fixed t, but in practice t will be small, and the limit $t \to 0$ corresponds to the classical problem in Section 2. We therefore consider various forms of asymptotic relations as $n \to \infty$, $h \to 0$, and $t \to 0$. There are three cases:

Case A. $h^{1/2}t^{-1} \to 0$, $u\{2\pi(hy)^{1/2}\, t^{-1}\} \to u(0) = (\pi/2)^{1/2}$, so

$$K_h(x) \sim (2a/t)\, K(x).$$

Thus $\operatorname{var} \hat{f}_1(x) = O(n^{-1}h^{-1}t^{-2})$ in terms of n, h, and t.

Case B. $h^{1/2}t^{-1} \to \infty$, $u\{2\pi(hy)^{1/2}\, t^{-1}\} \sim t(hy)^{-1/2}(2\pi)^{-1}$ by virtue of the relation $u(x) \sim x^{-1}$, $x \to \infty$. Thus in this case

$$K_h(x) \sim -(2/\pi^3)^{1/2}\, ah^{-1/2} K_1(x),$$

where K_1 is as in Section 2. Therefore, $\operatorname{var} \hat{f}_1(x) = O(n^{-1}h^{-2})$.

Case C. $2\pi h^{1/2}t^{-1} \to c$, $0 < c < \infty$. Then

$$K_h(x) \sim -(2/\pi)^{1/2}\, 2at^{-1} \int_0^{\infty} u(cy^{1/2})\, K'(y + x)\, dx.$$

We now have $\operatorname{var} \hat{f}_1(x) = O(n^{-1}h^{-1}t^{-2}) = O(n^{-1}h^{-2})$, where the constants of proportionality depend on c.

Case C is perhaps the most interesting case, since Cases A and B are limiting cases corresponding to $c \to 0$, $c \to \infty$, respectively. The broad conclusion is that similar results hold to those in the classical problem, with bias of $O(h^k)$, and variance of $O(n^{-1}h^{-2})$, but with constants of proportionality which also depend on t.

5. Conclusions

Although our results have been asymptotic in their character, there are some clear practical implications.

1. The optimal window width is of a larger order of magnitude for unfolding

problems than for ordinary density estimation. If $k=2$ (most kernels used in practice have $k=2$) then the optimal window width for the classical unfolding problem is $O(n^{-1/6})$, instead of $O(n^{-1/5})$ for ordinary density estimation.

2. If the optimal window width is used, then the mean squared error of the estimate is of a larger order of magnitude than for ordinary density estimation—$O(n^{-2/3})$ instead of $O(n^{-4/5})$ when $k=2$.

3. The results of Section 3 show that this increase of asymptotic mean squared error is inherent in the problem and is not a peculiarity of the kernel method of estimation.

4. Similar conclusions hold for other unfolding problems with explicit inversion formulae.

5. Taken together, the results show the statistical implications of the "ill-posed" nature of the problem. It is well known that solving Abel's integral equation is an ill-posed problem and that this creates difficulties for numerical solution. The kernel method *apparently* gets round these difficulties, provided a smooth kernel is used. The ill-posedness is still a problem, however, because of the more subtle statistical difficulty which we have highlighted in this paper.

APPENDIX 1: Statement of Assumptions

Assumptions on the kernel. We assume that the kernel function K satisfies (2.4) and

$$\int_{-\infty}^{\infty} |x|^k \{|K(x)| + |xK'(x)|\} \, dx < \infty$$

for some positive integer $k \geq 2$, and that both K and its derivative K' are bounded, continuous functions.

Assumptions on the density. Each of the formulae in Section 2 requires additional assumptions on the unknown density f, but we believe that they are all reasonable assumptions which should not inhibit the practical application of the method. For (2.5) we assume that f and hence f_1 are k times continuously differentiable in an interval containing x, and that $f(x)$ is bounded on $0<x<\infty$. For (2.10) we assume that $f, f', ..., f^{(k)}$ are all bounded on $0<x<\infty$ and

$$\int_{x}^{x+\varepsilon} (y-x)^{-1} \, |f^{(k)}(y) - f^{(k)}(x)| \, dy < \infty$$

for some $\varepsilon > 0$. Equations (2.8) and (2.11) both require that $f(x)$ is bounded on $0<x<\infty$. Equation (2.12) requires only that $f(x)$ is bounded on $0<x<\varepsilon$, for some $\varepsilon>0$.

APPENDIX 2: PROOF OF THEOREM

Fix x_0, B, and $\lambda > 0$. We shall construct densities $p_0, p_n \in C_k(B)$, with p_0 fixed and p_n depending on n, such that

$$p_0(x_0) - p_n(x_0) = \lambda n^{-k/(2k+2)}, \tag{A1}$$

and the corresponding folded densities q_0 and q_n, obtained via (1.4), satisfy

$$\int_0^\infty \{q_n(y) - q_0(y)\}^2 \{q_0(y)\}^{-1} \, dy = O(n^{-1}) \tag{A2}$$

Our argument is close to that of Farrell [4].

For $\delta \geqslant 0$, $k = 0, 1, 2, ...$, let $\Delta_{k\delta}$ denote a function satisfying the conditions:

(i) if $x \notin [-2^k\delta, 2^k\delta]$ then $\Delta_{k\delta}(x) = 0$,

(ii) $\Delta_{k\delta}(x)$ is an odd function of x,

(iii) the $(k-1)$th derivative of $\Delta_{k\delta}$ is absolutely continuous,

(iv) $\sup_{x \in \mathbb{R}} |\Delta_{k\delta}(x)| = 2^\gamma \delta^k$, where $\gamma = (k-1)(k-2)/2$,

(v) there exists a real sequence $\{c_k, k > 1\}$ such that

$$\int_{-\infty}^\infty \Delta_{k\delta}^2 (x) \, dx = c_k \delta^{(2k+1)},$$

(vi) $\Delta_{k\delta}(\delta x) \equiv \delta^k \Delta_k(x)$, where Δ_k is a function not depending on δ.

The existence of a function satisfying (i)–(v) is proved in Section 2a of Farrell [4] and Farrell's construction incidentally satisfies (vi) also. Let $p_0 \in C_k(B/2)$ have k continuous derivatives on $(0, \infty)$ and vanish outside a compact set and have the property $p_0(x) \equiv a > 0$ for $0 < x < x_0 + \varepsilon$, some $\varepsilon > 0$. Define $p_n(x) \equiv p_0(x) + \Delta_{k\delta}(x - x_0 - 2^{k-1}\delta)$, where $\delta = \delta(n) \to 0$ as $n \to \infty$. Set

$$m_i = \int_0^\infty x^{1/2} p_i(x) \, dx \qquad \text{and} \qquad q_i(y) = (2m_i)^{-1} \int_y^\infty (x-y)^{-1/2} p_i(x) \, dx.$$

Then $m_n q_n(y) - m_0 q_0(y)$ vanishes outside $(0, x_0 + \varepsilon/2)$, provided δ is sufficiently small. Furthermore, $q_0(y)$ is bounded away from zero uniformly in $y \in (0, x_0 + \varepsilon/2)$. Therefore for a constant $C > 0$,

$$\begin{aligned} J &\equiv \int_0^\infty \{m_n q_n(y) - m_0 q_0(y)\}^2 \{q_0(y)\}^{-1} \, dy \\ &\leqslant C \int_0^\infty dy \iint_{y < x_1 < x_2 < \infty} (x_1 - y)^{-1/2} (x_2 - y)^{-1/2} \Delta_{k\delta}(x_1 - x_0 - 2^{k-1}\delta) \\ &\quad \times \Delta_{k\delta}(x_2 - x_0 - 2^{k-1}\delta) \, dx_1 \, dx_2 \end{aligned}$$

$$= C\delta^{2k+2} \int_{-\infty}^{\infty} \Delta_k(z_1)\, dz \int_{z_1}^{\infty} \Delta_k(z_2)[2 \log\{(x_0 + \delta z_1 + 2^{k-1}\delta)^{1/2}$$

$$+ (x_0 + \delta z_2 + 2^{k-1}\delta)^{1/2}\} - \log(z_2 - z_1) - \log \delta]\, dz_2$$

$$= O(\delta^{2k+2}).$$

A simpler argument shows that $|m_n - m_0| = O(\delta^{k+2})$, and so the left-hand side of (A2) equals $O(\delta^{2k+2})$. Therefore (A2) holds if $\delta = O(n^{-1/(2k+2)})$.

There exists a positive constant c_k such that $\Delta_k(-2^{k-1}) = c_k$. If we take $\delta = dn^{-1/(2k+2)}$, where $d = (\lambda/c_k)^{1/k}$, then the left-hand side of (A1) equals $\Delta_{k\delta}(-2^{k-1}\delta) = \delta^k \Delta_k(-2^{k-1})$, which is the right-hand side. This proves (A1).

An immediate consequence of (A2) is that

$$E_{q0}[\{q_n(z_i)/q_0(z_i)\}^2] = 1 + \int_0^{\infty} \{q_n(x) - q_0(x)\}^2 \{q_0(x)\}^{-1}\, dx = 1 + O(n^{-1}).$$

Therefore,

$$P_{q_n}\{|\phi_n(Z_1, ..., Z_n) - p_n(x_0| \leqslant a_n\}$$

$$\leqslant [P_{q0}\{|\phi_n(Z_1, ..., Z_n) - p_n(x_0)| \leqslant a_n\}]^{1/2} \left[\prod_{i=1}^{n} E_{q0}\{q_n(Z_i)/q_0(Z_i)\}^2\right]^{1/2}$$

$$\leqslant C[P_{q0}\{|\phi_n(Z_1, ..., Z_n) - p_n(x_0)| \leqslant a_n\}]^{1/2}$$

for some constant C. Assumption (3.1) now implies that $P_{q0}\{|\phi_n(Z_1, ..., Z_n) - p_n(x_0)| \leqslant a_n\}$ is bounded away from zero, as well as that $P_{q0}\{|\phi_n(Z_1, ..., Z_n) - p_0(x_0)| \leqslant a_n\} \to 1$. Hence $|p_n(x_0) - p_0(x_0)| < 2a_n$ for large n, or equivalently, $a_n \geqslant (\lambda/2)\, n^{-k/(2k+2)}$. Since this is true for all $\lambda > 0$, we have (3.2).

ACKNOWLEDGMENT

The second author's work was supported by a Visiting Fellowship at the Australian National University, which is gratefully acknowledged.

REFERENCES

1. R. S. ANDERSSEN AND A. J. JAKEMAN, *J. Microscopy* **105**, 135 (1975).
2. M. S. BARTLETT, *Sankhya A* **25**, 245 (1963).
3. R. COLEMAN, *An Introduction to Mathematical Stereology* (Department of Theoretical Statistics, University of Aarhus, Denmark, 1979).
4. R. H. FARRELL, *Ann. Math. Statist.* **43**, 170 (1972).
5. M. J. FRYER, *J. Inst. Math. Appl.* **20**, 335 (1977).

6. A. J. JAKEMAN, *Utilitas Math.* **26**, 193 (1985).
7. A. J. JAKEMAN AND R. S. ANDERSSEN, *J. Microsc.* (*Oxford*) **105**, 121 (1975).
8. K. KANATANI AND O. ISHIKAWA, *J. Comput. Phys.* **57**, 229 (1985).
9. B. D. RIPLEY, *Spatial Statistics* (Wiley, New York, 1981).
10. L. A. SANTALÓ, *Trab. Estadist.* **6**, 181 (1955).
11. R. S. SINGH, *Biometrika* **66**, 177 (1979).
12. C. C. TAYLOR, *J. Microsc.* (*Oxford*) **132**, 57 (1983).
13. E. J. WEGMAN, *Technometrics* **14**, 533 (1972).