Asymptotic Normality in Density Support Estimation

Gérard BIAU [a,*], Benoît CADRE [b], David M. MASON [c] and Bruno PELLETIER [d]

[a] LSTA & LPMA
Université Pierre et Marie Curie – Paris VI
Boîte 158, 175 rue du Chevaleret
75013 Paris, France
gerard.biau@upmc.fr

[b] IRMAR, ENS Cachan Bretagne, CNRS, UEB
Campus de Ker Lann
Avenue Robert Schuman
35170 Bruz, France
Benoit.Cadre@bretagne.ens-cachan.fr

[c] University of Delaware
Food and Resource Economics
206 Townsend Hall
Newark, DE 19717, USA
davidm@udel.edu

[c] IRMAR, Université Rennes 2, CNRS, UEB
Campus Villejean
Place du Recteur Henri Le Moal, CS 24307
35043 Rennes Cedex, France
bruno.pelletier@uhb.fr

Abstract

Let $X_1, \ldots, X_n$ be n independent observations drawn from a multivariate probability density f with compact support S_f. This paper is devoted to the study of the estimator $\hat{S}_n$ of S_f defined as the union of balls centered at the X_i and with common radius r_n. Using tools from Riemannian geometry, and under mild assumptions on f and the sequence (r_n), we prove a central limit theorem for $\lambda(S_n \Delta S_f)$, where λ denotes the Lebesgue measure on $\mathbb{R}^d$ and Δ the symmetric difference operation.

Index Terms — Support estimation, Nonparametric statistics, Central limit theorem, Tubular neighborhood.

AMS 2000 Classification: 62G05, 62G20.

*Corresponding author.

1 Introduction

Let $X_1, \dots, X_n$ be independent and identically distributed observations drawn from an unknown probability density f defined on $\mathbb{R}^d$. It is assumed that $d \geq 2$ throughout this paper. We investigate the problem of estimating the support of f, i.e., the closed set

$$S_f = \overline{\{x \in \mathbb{R}^d : f(x) > 0\}},$$

based on the sample $X_1, \dots, X_n$. Here and elsewhere, $\overline{A}$ denotes the closure of a Borel set A. This problem is of interest due to the broad scope of its practical applications in applied statistics. These include medical diagnosis, machine condition monitoring, marketing and econometrics. For a review and a large list of references, we refer the reader to Molchanov (1998), Baíllo, Cuevas, and Justel (2000), Biau, Cadre, and Pelletier (2008) and Mason and Polonik (2009).

Devroye and Wise (1980) introduced the following very simple and intuitive estimator of S_f. It is defined as

$$S_n = \bigcup_{i=1}^{n} \mathcal{B}(X_i, r_n), \tag{1.1}$$

where $\mathcal{B}(x, r)$ denotes the closed Euclidean ball centered at x and of radius $r > 0$, and where (r_n) is an appropriately chosen sequence of positive smoothing parameters. For $x \in \mathbb{R}^d$, let

$$f_n(x) = \sum_{i=1}^{n} \mathbf{1}_{\mathcal{B}(x, r_n)}(X_i)$$

be the (unnormalized) kernel density estimator of f. We see that

$$S_n = \{x \in \mathbb{R}^d : f_n(x) > 0\}.$$

In other words, $S_n = S_{f_n}$, i.e., it is just a plug-in-type kernel estimator with kernel having a ball-shaped support. Baíllo, Cuevas, and Justel (2000) argue that this estimator is a good generalist when no *a priori* information is available about S_f. Moreover, from a practical perspective, the relative simplicity of the estimation strategy (1.1) is a major advantage over competing multidimensional set estimation techniques, which are often faced with a heavy computational burden.

Biau, Cadre, and Pelletier (2008) proved, under mild regularity assumptions on f and the sequence (r_n), that for an explicit constant $c > 0$,

$$\sqrt{nr_n^d}\,\mathbb{E}\lambda(S_n \Delta S_f) \to c,$$

where $\triangle$ denotes the symmetric difference operation and λ is the Lebesgue measure on $\mathbb{R}^d$. In the present paper, we go one step further and establish the asymptotic normality of $\lambda(S_n \triangle S_f)$. Precisely, our main Theorem 2.1 states, under appropriate regularity conditions on f and (r_n), that

$$\left(\frac{n}{r_n^d}\right)^{1/4}\Big(\lambda\left(S_n \triangle S_f\right) - \mathbb{E}\lambda\left(S_n \triangle S_f\right)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_f^2),$$

for some explicit positive variance σ_f^2.

Denoting by ∂S_f the boundary of S_f, it turns out that, under our conditions, $\lambda(\partial S_f) = 0$ and $f > 0$ on the interior of S_f. Therefore, we have the equality

$$\lambda\left(S_f \triangle \left\{x \in \mathbb{R}^d \,:\, f(x) > 0\right\}\right) = 0.$$

Thus, $\lambda(S_n \triangle S_f)$ may be expressed more conveniently as

$$\lambda(S_n \triangle S_f) = \int_{\mathbb{R}^d} \Big|\mathbf{1}\{f_n(x) > 0\} - \mathbf{1}\{f(x) > 0\}\Big| \mathrm{d}x.$$

This quantity is related to the so-called vacancy V_n left by randomly distributed spheres (see Hall 1985, 1988), which in this notation is

$$V_n = \lambda\left(S_f - S_n\right) = \int_{S_f} \Big|\mathbf{1}\{f_n(x) > 0\} - \mathbf{1}\{f(x) > 0\}\Big| \mathrm{d}x. \tag{1.2}$$

Hall (1985) has proved a number of central limit theorems for V_n. One of them, his Theorem 1, states that if f has support in $[0,1]^d$ and is continuous then, as long as $nr_n^d \to a$ where $0 < a < \infty$, for some $0 < \sigma_a^2 < \infty$,

$$\sqrt{n}\left(V_n - \mathbb{E}V_n\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_a^2).$$

As pointed out in Hall's paper, and to the best of our knowledge, the case when $nr_n^d \to \infty$ has not been examined, except for some restricted cases in dimension 1. It turns out that, by adapting our arguments to the vacancy problem, we are also able to prove a general central limit theorem for V_n when $nr_n^d \to \infty$, thereby extending Hall's results. For more about large sample properties of vacancy and their applications consult Chapter 3 of Hall

(1988).

Another result closely related to ours is the following special case of the main theorem in Mason and Polonik (2009). For any $0 < c < \sup\{f(x) : x \in \mathbb{R}^d\}$, let $C(c) = \{x \in \mathbb{R}^d : f(x) > c\}$ and $\hat{C}_n(c) = \{x \in \mathbb{R}^d : \hat{f}_n(x) > c\}$, where $\hat{f}_n$ denotes a kernel estimator of f. Then

$$\lambda\left(\hat{C}_n(c) \triangle C(c)\right) = \int_{\mathbb{R}^d} \left| \mathbf{1}\{\hat{f}_n(x) > c\} - \mathbf{1}\{f(x) > c\} \right| \mathrm{d}x.$$

Mason and Polonik (2009) prove, subject to regularity conditions on f, as long as $\sqrt{nr_n^{d+2}} \to \gamma$, with $0 \le \gamma < \infty$ and $nr_n^d / \log n \to \infty$, where $\gamma = 0$ in the case $d = 1$, that for some $0 < \sigma_c^2 < \infty$,

$$\left(\frac{n}{r_n^d}\right)^{1/4} \left(\lambda\left(\hat{C}_n(c) \triangle C(c)\right) - \mathbb{E}\lambda\left(\hat{C}_n(c) \triangle C(c)\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_c^2).$$

The paper is organized as follows. In Section 2, we first set out notation and assumptions, and then state our main results. Section 3 is devoted to the proofs.

2 Asymptotic normality of $\lambda(S_n \Delta S_f)$

2.1 Notation and assumptions

Throughout the paper, we shall impose the following set of assumptions related to the support of f. To state some of them we shall require a few concepts and terms from Riemannian geometry. For a good introduction to the subject we refer the reader, for instance, to the book by Gallot, Hulin and Lafontaine (2004). We make use of the notation $\overset{\circ}{S}_f$ to denote the interior of S_f.

Assumption Set 1

(a) The support S_f of f is compact in $\mathbb{R}^d$, with $d \ge 2$.

(b) f is of class $\mathcal{C}^1$ on $\mathbb{R}^d$, and of class $\mathcal{C}^2$ on $\mathcal{T} \cap \overset{\circ}{S}_f$, where $\mathcal{T}$ is a tubular neighborhood of S_f.

(c) The boundary ∂S_f of S_f is a smooth submanifold of $\mathbb{R}^d$ of codimension 1.

(d) The set $\{x \in \mathbb{R}^d : f(x) > 0\}$ is connected.

(e) $f > 0$ on $\overset{\circ}{S}_f$.

Roughly speaking, under Assumption 1-(c), the boundary ∂S_f is a subset of dimension $(d-1)$ of the ambient space $\mathbb{R}^d$. For instance, consider a density supported on the unit ball of $\mathbb{R}^d$. In this case, the boundary is the unit sphere of dimension $(d-1)$. Note also that one can relax Assumption 1-(d) to the case where the set $\{x \in \mathbb{R}^d : f(x) > 0\}$ has multiple connected components, see Remark 3.3 in Biau, Cadre, and Pelletier (2008).

More precisely, under Assumption 1-(c), ∂S_f is a smooth Riemannian submanifold with Riemannian metric, denoted by σ, induced by the canonical embedding of ∂S_f in $\mathbb{R}^d$. The volume measure on $(\partial S_f, \sigma)$ will be denoted by v_σ. Furthermore, $(\partial S_f, \sigma)$ is compact and without boundary. Then by the tubular neighborhood theorem (see e.g., Gray, 1990; Bredon, 1993, p. 93), ∂S_f admits a tubular neighborhood of radius $\rho > 0$,

$$\mathcal{V}(\partial S_f, \rho) = \left\{x \in \mathbb{R}^d : \operatorname{dist}(x, \partial S_f) < \rho\right\},$$

i.e., each point $x \in \mathcal{V}(\partial S_f, \rho)$ projects uniquely onto ∂S_f. Let $\{e_p\,;\, p \in \partial S_f\}$ be the unit-norm section of the normal bundle $T\partial S_f^\perp$ that is pointing inwards, i.e., for all $p \in \partial S_f$, e_p is the unit normal vector to ∂S_f directed towards the interior of S_f. Then each point $x \in \mathcal{V}(\partial S_f, \rho)$ may be expressed as

$$x = p + ve_p, \tag{2.1}$$

where $p \in \partial S_f$, and where $v \in \mathbb{R}$ satisfies $|v| \leq \rho$. Moreover, given a Lebesgue integrable function φ on $\mathcal{V}(\partial S_f, \rho)$, we may write

$$\int_{\mathcal{V}(\partial S_f,\rho)} \varphi(x)\mathrm{d}x = \int_{\partial S_f} \int_{-\rho}^{\rho} \varphi(p + ve_p)\Theta(p, u)\mathrm{d}u\, v_\sigma(\mathrm{d}p), \tag{2.2}$$

where Θ is a $\mathcal{C}^\infty$ function satisfying $\Theta(p, 0) = 1$ for all $p \in \partial S_f$. (See Appendix B in Biau, Cadre, and Pelletier, 2008.)

Denote by $D^2_{e_p}$ the directional differentiation operator of order 2 on $\mathcal{V}(\partial S_f, \rho)$ in the direction e_p. It will be seen in the proofs in Section 3 that the variance in our central limit theorem is determined by the second order behavior of f near the boundary of its support. Therefore to derive this variance we shall need the following set of second order smoothness assumptions on f.

Assumption Set 2

(a) There exists $\rho > 0$ such that, for all $p \in \partial S_f$, the map $u \mapsto f(p + ue_p)$ is of class $\mathcal{C}^2$ on $[0, \rho]$.

(b) There exists $\rho > 0$ such that

$$0 < \sup_{p \in \partial S_f} \sup_{0 \leq u \leq \rho} D^2_{e_p} f(p + u e_p) < \infty.$$

(c) There exists $\rho > 0$ such that

$$\inf_{p \in \partial S_f} \inf_{0 \leq u \leq \rho} D^2_{e_p} f(p + u e_p) > 0.$$

For similar smoothness assumptions see Section 2.4 of Mason and Polonik (2009). The imposition of such conditions appears to be unavoidable to derive a central limit theorem. Note also that Assumption Sets 1 and 2 are the same as the ones used in Biau, Cadre, and Pelletier (2008). In particular, we assume throughout that the density f is *continuous* on $\mathbb{R}^d$. Thus, we are in the case of a non-sharp boundary, i.e., f decreases continuously to zero at the boundary of its support. The case where f has sharp boundary requires a different approach (see for example Härdle, Park, and Tsybakov, 1995). The analytical assumptions on f (Assumption Set 2) are stipulations on the local behavior of f at the boundary of the support. In particular, the restrictions on f imply that inside the support and close to the boundary the maps $u \mapsto f(p + u e_p)$, with $p \in \partial S_f$, are strictly convex (see the Appendix).

2.2 Main result

Let

$$\sigma_f^2 = 2^d \int_{\partial S_f} \int_0^\infty \int_{\mathcal{B}(0,1)} \Phi(p, t, u) \mathrm{d}u \mathrm{d}t v_\sigma(\mathrm{d}p), \tag{2.3}$$

with

$$\Phi(p, t, u) = \exp\left(-\omega_d D^2_{e_p} f(p) t^2\right) \left[\exp\left(\beta(u) D^2_{e_p} f(p) \frac{t^2}{2}\right) - 1\right],$$

ω_d denoting the volume of $\mathcal{B}(0,1)$ and

$$\beta(u) = \lambda\left(\mathcal{B}(0,1) \cap \mathcal{B}(2u, 1)\right).$$

Remark 2.1 *Let Γ be the Gamma function. We note that $\beta(u)$ has the closed expression (Hall, 1988, p. 23)*

$$\beta(u) = \begin{cases} \dfrac{2\pi^{(d-1)/2}}{\Gamma\left(\frac{1}{2} + \frac{d}{2}\right)} \displaystyle\int_{|u|}^1 (1 - y^2)^{(d-1)/2} \mathrm{d}y, & \text{if } 0 \leq |u| \leq 1 \\ 0, & \text{if } |u| > 1, \end{cases}$$

which, in particular, gives

$$\beta(0) = \omega_d = \frac{\pi^{d/2}}{\Gamma\left(1 + \frac{d}{2}\right)}.$$

We are now ready to state our main result.

Theorem 2.1 *Suppose that both Assumption Sets 1 and 2 are satisfied. If (r.i) $r_n \to 0$, (r.ii) $nr_n^d/(\ln n)^{4/3} \to \infty$ and (r.iii) $nr_n^{d+1} \to 0$, then*

$$\left(\frac{n}{r_n^d}\right)^{1/4} \Big(\lambda(S_n \triangle S_f) - \mathbb{E}\lambda(S_n \triangle S_f)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_f^2),$$

where $\sigma_f^2 > 0$ is as in (2.3).

Remark 2.2 *A referee pointed out that the methods in the paper may be applicable to obtain a central limit theorem for the histogram-based support estimator studied in Baíllo and Cuevas (2006). The reference Cuevas, Fraiman, and Rodríguez-Casal (2007) should be a starting point for such an investigation.*

It is known from Cuevas and Rodríguez-Casal (2004) that the choice $r_n = \mathrm{O}((\ln n/n)^{1/d})$ gives the fastest convergence rate of S_n towards S_f for the Hausdorff metric, that is $\mathrm{O}((\ln n/n)^{1/d})$. For such a radius choice, the concentration speed of $\lambda(S_n \triangle S_f)$ around its expectation as given by Theorem 2.1 is $\mathrm{O}(\sqrt{n}/(\ln n)^{1/4})$, close to the parametric rate.

Theorem 2.1 assumes $d \geq 2$ (Assumption 1-(a)). We restrict ourselves to the case $d \geq 2$ for the sake of technical simplicity. However, the case $d = 1$ can be derived with minor adaptations, assuming $r_n \to 0$, $nr_n/(\ln n)^{4/3} \to \infty$, and $nr_n^{3/2} \to 0$. In fact, the one-dimensional setting has already been explored in the related context of vacancy estimation (Hall, 1984).

As we mentioned in the introduction, the quantity $\lambda(S_n \triangle S_f)$ is closely related to the vacancy V_n (Hall 1985, 1988), which is defined as in (1.2). A close inspection of the proof of Theorem 2.1 reveals that taking intersection with S_f in the integrals does not effect things too much and, in fact, the asymptotic distributional behaviors of $\lambda(S_n \triangle S_f)$ and V_n are nearly identical. As a consequence, we obtain the following result:

Theorem 2.2 *Suppose that both Assumption Sets 1 and 2 are satisfied. If (r.i), (r.ii) and (r.iii) hold, then*

$$\left(\frac{n}{r_n^d}\right)^{1/4} (V_n - \mathbb{E}V_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_f^2),$$

where $\sigma_f^2 > 0$ is as in (2.3).

Surprisingly, the limiting variance σ_f^2 remains as in (2.3). Theorem 2.2 was motivated by a remark by Hall (1985), who pointed out that a central limit theorem for vacancy in the case $nr_n^d \to \infty$ remained open.

3 Proof of Theorem 2.1

Our proof of Theorem 2.1 will borrow elements from Mason and Polonik (2009).

Set

$$\varepsilon_n = \frac{1}{(nr_n^d)^{1/4}}. \tag{3.1}$$

Observe that, from (r.ii) and (r.iii), the sequence (ε_n) satisfies (e.i) $\varepsilon_n \to 0$ and (e.ii) $\varepsilon_n \sqrt{nr_n^d} \to \infty$. For future reference we note that from (r.i) and (r.iii), we get that

$$\frac{r_n}{\varepsilon_n} \to 0. \tag{3.2}$$

Set

$$\mathcal{E}_n = \{x \in \mathbb{R}^d : f(x) \le \varepsilon_n\}.$$

Furthermore, let

$$L_n(\varepsilon_n) = \int_{\mathcal{E}_n} \Big| \mathbf{1}\{f_n(x) > 0\} - \mathbf{1}\{f(x) > 0\} \Big| \mathrm{d}x$$

and

$$\overline{L}_n(\varepsilon_n) = \int_{\mathcal{E}_n^c} \Big| \mathbf{1}\{f_n(x) > 0\} - \mathbf{1}\{f(x) > 0\} \Big| \mathrm{d}x.$$

Noting that, under Assumption Set 1, $\lambda(S_n \Delta S_f) = L_n(\varepsilon_n) + \overline{L}_n(\varepsilon_n)$, our plan is to show that

$$\left(\frac{n}{r_n^d}\right)^{1/4} \Big(L_n(\varepsilon_n) - \mathbb{E}L_n(\varepsilon_n) \Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_f^2) \tag{3.3}$$

and

$$\left(\frac{n}{r_n^d}\right)^{1/4} \Big(\overline{L}_n(\varepsilon_n) - \mathbb{E}\overline{L}_n(\varepsilon_n) \Big) \xrightarrow{\mathbb{P}} 0, \tag{3.4}$$

which together imply the statement of Theorem 2.1. To prove a central limit theorem for the random variable $L_n(\varepsilon_n)$, it turns out to be more convenient to first establish one for the Poissonized version of it formed by replacing $f_n(x)$ with

$$\pi_n(x) = \sum_{i=1}^{N_n} \mathbf{1}_{\mathcal{B}(x,r_n)}(X_i),$$

where N_n is a mean n Poisson random variable independent of the sample $X_1, \dots, X_n$. By convention, we set $\pi_n(x) = 0$ whenever $N_n = 0$. The Poissonized version of $L_n(\varepsilon_n)$ is then defined by

$$\Pi_n(\varepsilon_n) = \int_{\mathcal{E}_n} \Big| \mathbf{1}\{\pi_n(x) > 0\} - \mathbf{1}\{f(x) > 0\} \Big| \mathrm{d}x.$$

The proof of Theorem 2.1 is organized as follows. First (Subsection 3.1), we determine the exact asymptotic behavior of the variance of $\Pi_n(\varepsilon_n)$. Then (Subsection 3.2), we prove a central limit theorem for $\Pi_n(\varepsilon_n)$. By means of a de-Poissonization result (Subsection 3.3), we then infer (3.3). In a final step (Subsection 3.4) we prove (3.4), which completes the proof of Theorem 2.1. This Poissonization/de-Poissonization methodology goes back to at least Beirlant, Györfi, and Lugosi (1994).

3.1 Exact asymptotic behavior of Var$(\Pi_n(\varepsilon_n))$

Let

$$\Delta_n(x) = \Big| \mathbf{1}\{\pi_n(x) > 0\} - \mathbf{1}\{f(x) > 0\} \Big|.$$

In the sequel, the letter C will denote a positive constant, the value of which may vary from line to line.

Let (ε_n) be the sequence of positive real numbers defined in (3.1). In this subsection, we intend to prove that, under the conditions of Theorem 2.1,

$$\lim_{n\to\infty} \sqrt{\frac{n}{r_n^d}}\, \mathrm{Var}\left(\Pi_n(\varepsilon_n)\right) = \sigma_f^2, \tag{3.5}$$

where σ_f^2 is as in (2.3).

Towards this goal, observe first that

$$\Pi_n(\varepsilon_n) = \int_{\tilde{\mathcal{E}}_n} \Big| \mathbf{1}\{\pi_n(x) > 0\} - \mathbf{1}\{f(x) > 0\} \Big| \mathrm{d}x,$$

where we set

$$\tilde{\mathcal{E}}_n = \mathcal{E}_n \cap S_f^{r_n},$$

with

$$S_f^{r_n} = \left\{x \in \mathbb{R}^d \,:\, \mathrm{dist}(x, S_f) \le r_n\right\}.$$

Clearly,

$$\mathrm{Var}(\Pi_n\left(\varepsilon_n\right)) = \int_{\tilde{\mathcal{E}}_n} \int_{\tilde{\mathcal{E}}_n} \mathbb{C}\left(\Delta_n(x), \Delta_n(y)\right) \mathrm{d}x \mathrm{d}y,$$

where here and elsewhere $\mathbb{C}$ denotes 'covariance'. Since $\Delta_n(x)$ and $\Delta_n(y)$ are independent whenever $\|x-y\| > 2r_n$, we may write

$$\mathrm{Var}\left(\Pi_n(\varepsilon_n)\right) = \int_{\tilde{\mathcal{E}}_n} \int_{\tilde{\mathcal{E}}_n} \mathbf{1}\left\{\|x-y\| \le 2r_n\right\} \mathbb{C}\left(\Delta_n(x), \Delta_n(y)\right) \mathrm{d}x \mathrm{d}y.$$

Using the change of variable $y = x + 2r_n u$, we obtain

$$\begin{aligned}&\mathrm{Var}(\Pi_n\left(\varepsilon_n\right))\\ &\quad= 2^d r_n^d \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{\tilde{\mathcal{E}}_n}(x) \mathbf{1}_{\tilde{\mathcal{E}}_n}(x + 2r_n u) \mathbf{1}_{\mathcal{B}(0,1)}(u) \mathbb{C}\left(\Delta_n(x), \Delta_n(x + 2r_n u)\right) \mathrm{d}x \mathrm{d}u.\end{aligned}$$

By construction, whenever n is large enough, $\tilde{\mathcal{E}}_n$ is included in the tubular neighborhood $\mathcal{V}(\partial S_f, \rho)$ of ∂S_f of radius $\rho > 0$. In this case, each $x \in \tilde{\mathcal{E}}_n$ may be written as $x = p + v e_p$ as described in (2.1). Hence, for all large enough n, we obtain

$$\begin{aligned}&\mathrm{Var}(\Pi_n\left(\varepsilon_n\right))\\ &\quad= 2^d r_n^d \int_{\partial S_f} \int_{-r_n}^{\rho} \int_{\mathcal{B}(0,1)} \mathbf{1}_{\tilde{\mathcal{E}}_n}(p + v e_p) \mathbf{1}_{\tilde{\mathcal{E}}_n}(p + v e_p + 2r_n u)\\ &\qquad \times \Theta(p, v) \mathbb{C}\left(\Delta_n(p + v e_p), \Delta_n(p + v e_p + 2r_n u)\right) \mathrm{d}u \mathrm{d}v v_\sigma(\mathrm{d}p).\end{aligned}$$

For all $p \in \partial S_f$, let $\kappa_p(\varepsilon_n)$ be the distance between p and the point x of the set $\{x \in \mathbb{R}^d : f(x) = \varepsilon_n\}$ such that the vector $x - p$ is orthogonal to ∂S_f. Using the change of variable $v = t/\sqrt{n r_n^d}$, we may write

$$\begin{aligned}&\mathrm{Var}\left(\Pi_n(\varepsilon_n)\right)\\ &\quad= \frac{2^d r_n^d}{\sqrt{n r_n^d}} \int_{\partial S_f} \int_{-\sqrt{n r_n^{d+2}}}^{\sqrt{n r_n^d}\kappa_p(\varepsilon_n)} \int_{\mathcal{B}(0,1)} \mathbf{1}_{\tilde{\mathcal{E}}_n}\left(p + \frac{t}{\sqrt{n r_n^d}} e_p + 2r_n u\right) \Theta\left(p, \frac{t}{\sqrt{n r_n^d}}\right)\\ &\qquad \times \mathbb{C}\left(\Delta_n\left(p + \frac{t}{\sqrt{n r_n^d}} e_p\right), \Delta_n\left(p + \frac{t}{\sqrt{n r_n^d}} e_p + 2r_n u\right)\right) \mathrm{d}u \mathrm{d}t v_\sigma(\mathrm{d}p).\end{aligned}$$

For a justification of this change of variable, refer to equation (2.2) and equation (4.2) in the Appendix. By conditions (r.i) and (r.iii), $n r_n^{d+2} \to 0$. Consequently,

$$\sqrt{\frac{n}{r_n^d}}\, \mathrm{Var}\left(\Pi_n(\varepsilon_n)\right)$$

$$
\begin{aligned}
&= \mathrm{o}(1) \\
&\quad + 2^d \int_{\partial S_f} \int_0^{\sqrt{nr_n^d}\kappa_p(\varepsilon_n)} \int_{\mathcal{B}(0,1)} \mathbf{1}_{\tilde{\mathcal{E}}_n}\left(p + \frac{t}{\sqrt{nr_n^d}} s e_p + 2r_n u\right) \Theta\left(p, \frac{t}{\sqrt{nr_n^d}}\right) \\
&\quad \times \mathbb{C}\left(\Delta_n\left(p + \frac{t}{\sqrt{nr_n^d}} e_p\right), \Delta_n\left(p + \frac{t}{\sqrt{nr_n^d}} e_p + 2r_n u\right)\right) \mathrm{d}u \mathrm{d}t v_\sigma(\mathrm{d}p).
\end{aligned} \tag{3.6}
$$

To get the limit as $n \to \infty$ of the above integral, we will need the following lemma, whose proof is deferred to the end of the subsection.

Lemma 3.1 *Let $p \in \partial S_f$, $t > 0$ and $u \in \mathcal{B}(0,1)$ be fixed. Suppose that the conditions of Theorem 2.1 hold. Then*

$$
\lim_{n\to\infty} \mathbb{C}\left(\Delta_n\left(p + \frac{t}{\sqrt{nr_n^d}} e_p\right), \Delta_n\left(p + \frac{t}{\sqrt{nr_n^d}} e_p + 2r_n u\right)\right) = \Phi(p,t,u),
$$

where $\Phi(p,t,u)$ is defined in Theorem 2.1.

Returning to the proof of (3.5), we notice that by (4.4) in the Appendix and (e.ii) we have $\sqrt{nr_n^d}\kappa_p(\varepsilon_n) \to \infty$ as $n \to \infty$ and $\Theta(p,0) = 1$. Therefore, using Lemma 3.1 and the fact that for all $t > 0$ and $u \in \mathcal{B}(0,1)$

$$
\mathbf{1}_{\tilde{\mathcal{E}}_n}\left(p + \frac{t}{\sqrt{nr_n^d}} + 2r_n u\right) \to 1 \quad \text{as } n \to \infty,
$$

we conclude that the function inside the integral in (3.6) converges pointwise to $\Phi(p,t,u)$ as $n \to \infty$.

We now proceed to sufficiently bound the function inside the integral in (3.6) to be able to apply the Lebesgue dominated convergence theorem. Towards this goal, fix $p \in \partial S_f$, $u \in \mathcal{B}(0,1)$ and $0 < t \leq \sqrt{nr_n^d}\kappa_p(\varepsilon_n)$. Since $\Delta_n(x) \leq 1$ for all $x \in \mathbb{R}^d$, using the inequality $|\mathbb{C}(Y_1,Y_2)| \leq 2\mathbb{E}|Y_1|$ whenever $|Y_2| \leq 1$, we have

$$
\begin{aligned}
&\left|\mathbb{C}\left(\Delta_n\left(p + \frac{t}{\sqrt{nr_n^d}} e_p\right), \Delta_n\left(p + \frac{t}{\sqrt{nr_n^d}} e_p + 2r_n u\right)\right)\right| \\
&\quad \leq 2\,\mathbb{E}\,\Delta_n\left(p + \frac{t}{\sqrt{nr_n^d}} e_p\right).
\end{aligned} \tag{3.7}
$$

By the bound in (4.3) in the Appendix, we see that

$$\sup_{p\in\partial S_f} \kappa_p(\varepsilon_n) \to 0 \quad \text{as} \quad n\to\infty. \tag{3.8}$$

Then, since e_p is a normal vector to ∂S_f at p which is directed towards the interior of S_f, there exists an integer N_0 independent of p, t and u such that, for all $n \geq N_0$, the point $p + (t/\sqrt{nr_n^d})e_p$ belongs to the interior of S_f. Therefore, $f(p + (t/\sqrt{nr_n^d})e_p) > 0$ and, letting

$$\varphi_n(x) = \mathbb{P}\left(X \in \mathcal{B}(x, r_n)\right),$$

we obtain

$$\begin{aligned}
\mathbb{E}\,\Delta_n\left(p + \frac{t}{\sqrt{nr_n^d}}e_p\right) &= \mathbb{P}\left(\pi_n\left(p + \frac{t}{\sqrt{nr_n^d}}e_p\right) = 0\right)\\
&= \mathbb{E}\left[\mathbb{P}\left(\forall i \leq N_n \,:\, X_i \notin \mathcal{B}\left(p + \frac{t}{\sqrt{nr_n^d}}e_p, r_n\right)\middle| N_n\right)\right]\\
&= \mathbb{E}\left[1 - \varphi_n\left(p + \frac{t}{\sqrt{nr_n^d}}e_p\right)\right]^{N_n}\\
&= \exp\left[-n\varphi_n\left(p + \frac{t}{\sqrt{nr_n^d}}e_p\right)\right], \qquad (3.9)
\end{aligned}$$

where we used the fact that N_n is a mean n Poisson distributed random variable independent of the sample. By a slight adaptation of the proof of Lemma A.1 in Biau, Cadre, and Pelletier (2008) and under Assumption 1-(b), one deduces that for all $x \in \mathcal{V}(\partial S_f, \rho) \cap \overset{\circ}{S}_f$, there exists a quantity $K_n(x)$ such that

$$\varphi_n(x) = r_n^d \omega_d f(x) + r_n^{d+2} K_n(x) \quad \text{and} \quad \sup_n \sup_{x\in\mathcal{V}(\partial S_f,\rho)\cap\overset{\circ}{S}_f} |K_n(x)| < \infty. \tag{3.10}$$

For all x in $\mathcal{V}(\partial S_f, \rho)$ written as $x = p + ue_p$ with $p \in \partial S_f$ and $0 \leq u \leq \rho$, a Taylor expansion of f at p gives the expression

$$f(x) = \frac{1}{2} D_{e_p}^2 f(p + \xi e_p) u^2,$$

for some $0 \leq \xi \leq u$ since, by Assumption 1-(b), $D_{e_p} f(p) = 0$. Thus, in our context, expanding f at p, we may write

$$n\varphi_n\left(p + \frac{t}{\sqrt{nr_n^d}}e_p\right) = \omega_d D_{e_p}^2 f(p + \xi e_p)\frac{t^2}{2} + nr_n^{d+2} R_n(p, t),$$

for some $0 \le \xi \le \kappa_p(\varepsilon_n)$, and where $R_n(p,t)$ satisfies

$$\sup_n \sup \left\{ |R_n(p,t)| : p \in \partial S_f \text{ and } 0 \le t \le \sqrt{nr_n^d}\kappa_p(\varepsilon_n) \right\} < \infty.$$

Furthermore, by (3.8), each point $p + \xi e_p$ falls in the tubular neighborhood $\mathcal{V}(\partial S_f, \rho)$ for all large enough n. Consequently, by Assumption 2-(c) there exists $\alpha > 0$ independent of n and $N_1 \ge N_0$ independent of p, t and u such that, for all $n \ge N_1$,

$$\inf_{p \in \partial S_f} D_{e_p}^2 f(p + \xi e_p) > 2\alpha.$$

This, together with identity (3.9) and (r.i), (r.iii), which imply $nr_n^{d+2} \to 0$, leads to

$$\mathbb{E}\, \Delta_n \left(p + \frac{t}{\sqrt{nr_n^d}} e_p \right) \le C \exp(-\omega_d \alpha t^2) \tag{3.11}$$

for $n \ge N_1$ and all

$$0 \le t \le \sqrt{nr_n^d} \sup_{p \in \partial S_f} \kappa_p(\varepsilon_n).$$

Thus, using inequality (3.11), we deduce that the function on the left hand side of (3.7) is dominated by an integrable function of (p,t,u), which is independent of n provided $n \ge N_1$. Finally, we are in a position to apply the Lebesgue dominated convergence theorem, to conclude that

$$\lim_{n\to\infty} \sqrt{\frac{n}{r_n^d}}\, \mathrm{Var}\left(\Pi_n(\varepsilon_n)\right) = 2^d \int_{\partial S_f} \int_0^\infty \int_{\mathcal{B}(0,1)} \Phi(p,t,u) \mathrm{d}u \mathrm{d}t v_\sigma(\mathrm{d}p) = \sigma_f^2.$$

To be complete, it remains to prove Lemma 3.1.

Proof of Lemma 3.1 Let $x_n = p + (t/\sqrt{nr_n^d})e_p$. Since $nr_n^d \to \infty$ and $nr_n^{d+2} \to 0$, both x_n and $x_n + 2r_n u$ lie in the interior of S_f for all large enough n. As a consequence, $f(x_n) > 0$ and $f(x_n + 2r_n u) > 0$ for all large enough n. Thus,

$$\begin{aligned}
&\mathbb{C}\left(\Delta_n(x_n), \Delta_n(x_n + 2r_n u)\right) \\
&\quad = \mathbb{C}\left(\mathbf{1}\{\pi_n(x_n) = 0\}, \mathbf{1}\{\pi_n(x_n + 2r_n u) = 0\}\right) \\
&\quad = \mathbb{P}\left(\pi_n(x_n) = 0, \pi_n(x_n + 2r_n u) = 0\right) - \mathbb{P}\left(\pi_n(x_n) = 0\right) \mathbb{P}\left(\pi_n(x_n + 2r_n u) = 0\right) \\
&\quad = \mathbb{P}\left(\forall i \le N_n : X_i \notin \mathcal{B}(x_n, r_n) \cup \mathcal{B}(x_n + 2r_n u, r_n)\right) \\
&\qquad - \mathbb{P}\left(\forall i \le N_n : X_i \notin \mathcal{B}(x_n, r_n)\right) \mathbb{P}\left(\forall i \le N_n : X_i \notin \mathcal{B}(x_n + 2r_n u, r_n)\right) \\
&\quad = \exp\left[-n\mu\left(\mathcal{B}(x_n, r_n) \cup \mathcal{B}(x_n + 2r_n u, r_n)\right)\right] \\
&\qquad - \exp\left[-n\mu(\mathcal{B}(x_n, r_n)) - n\mu(\mathcal{B}(x_n + 2r_n u, r_n))\right],
\end{aligned}$$

where μ denotes the distribution of X. Let $B_n = \mathcal{B}(x_n, r_n) \cap \mathcal{B}(x_n + 2r_n u, r_n)$. Using the equality

$$\mu\left(\mathcal{B}(x_n, r_n) \cup \mathcal{B}(x_n + 2r_n u, r_n)\right) = \varphi_n(x_n) + \varphi_n(x_n + 2r_n u) - \mu(B_n),$$

we obtain

$$\begin{aligned} &\mathbb{C}\left(\Delta_n(x_n), \Delta_n(x_n + 2r_n u)\right) \\ &\quad = \exp\left[-n\left(\varphi_n(x_n) + \varphi_n(x_n + 2r_n u)\right)\right]\left[\exp\left(n\mu(B_n)\right) - 1\right]. \end{aligned} \tag{3.12}$$

Now, $\mu(B_n)$ may be expressed as

$$\mu(B_n) = f(x_n)\lambda(B_n) + \int_{B_n} \left(f(v) - f(x_n)\right) \mathrm{d}v.$$

Since f is of class $\mathcal{C}^1$ on $\mathbb{R}^d$, by developing f at x_n in the above integral, we obtain

$$\int_{B_n} \left(f(v) - f(x_n)\right) \mathrm{d}v = r_n^{d+1} R_n,$$

where R_n satisfies

$$|R_n| \leq C \sup_K \|\mathrm{grad}\, f\|,$$

and K is some compact subset of $\mathbb{R}^d$ containing ∂S_f and of nonempty interior. Next, note that $\lambda(B_n) = r_n^d \beta(u)$, where

$$\beta(u) = \lambda\left(\mathcal{B}(0,1) \cap \mathcal{B}(2u, 1)\right).$$

Therefore, expanding f at p in the direction e_p, we obtain

$$\mu(B_n) = \beta(u)\frac{t^2}{2n} D^2_{e_p} f\Big(p + \xi \frac{t}{\sqrt{n r_n^d}} e_p\Big) + r_n^{d+1} R_n,$$

where $\xi \in (0,1)$. Hence by (r.iii),

$$\lim_{n\to\infty} n\mu(B_n) = \beta(u) D^2_{e_p} f(p) \frac{t^2}{2}.$$

The above limit, together with identity (3.12) and (3.10), leads to the desired result. $\square$

3.2 Central limit theorem for $\Pi_n(\varepsilon_n)$

In this subsection we establish a central limit theorem for $\Pi_n(\varepsilon_n)$. Set

$$S_n(\varepsilon_n) = \frac{a_n\left(\Pi_n(\varepsilon_n) - \mathbb{E}\Pi_n(\varepsilon_n)\right)}{\sigma_n},$$

where $a_n = (n/r_n^d)^{1/4}$ and

$$\sigma_n^2 = \mathrm{Var}\Big(a_n\left(\Pi_n(\varepsilon_n) - \mathbb{E}\Pi_n(\varepsilon_n)\right)\Big).$$

We shall verify that as $n \to \infty$

$$S_n(\varepsilon_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.13}$$

To show this we require the following special case of Theorem 1 of Shergin (1990).

Fact 3.1 *Let $(X_{\mathbf{i},n} : \mathbf{i} \in \mathbb{Z}^d)$ denote a triangular array of mean zero m-dependent random fields, and let $\mathcal{J}_n \subset \mathbb{Z}^d$ be such that*

(i) $\mathrm{Var}\left(\sum_{\mathbf{i}\in\mathcal{J}_n} X_{\mathbf{i},n}\right) \to 1$ *as $n \to \infty$, and*

(ii) For some $2 < s < 3$, $\sum_{\mathbf{i}\in\mathcal{J}_n} \mathbb{E}|X_{\mathbf{i},n}|^s \to 0$ as $n \to \infty$.

Then

$$\sum_{\mathbf{i}\in\mathcal{J}_n} X_{\mathbf{i},n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

We use Shergin's result as follows. Recall the definition of ε_n in (3.1) and also that

$$\mathrm{Var}(\Pi_n(\varepsilon_n)) = \int_{\tilde{\mathcal{E}}_n}\int_{\tilde{\mathcal{E}}_n} \mathbb{C}\left(\Delta_n(x), \Delta_n(y)\right) \mathrm{d}x\mathrm{d}y,$$

with

$$\tilde{\mathcal{E}}_n = \mathcal{E}_n \cap S_f^{r_n}.$$

Next, consider the regular grid given by

$$A_{\mathbf{i}} = (x_{i_1}, x_{i_1+1}] \times \ldots \times (x_{i_d}, x_{i_d+1}],$$

where $\mathbf{i} = (i_1, \ldots, i_d)$, $i_1, \ldots, i_d \in \mathbb{Z}$ and $x_i = i\, r_n$ for $i \in \mathbb{Z}$. Define

$$R_{\mathbf{i}} = A_{\mathbf{i}} \cap \tilde{\mathcal{E}}_n.$$

With $\mathcal{J}_n = \{\mathbf{i} \in \mathbb{Z}^d : A_{\mathbf{i}} \cap \tilde{\mathcal{E}}_n \neq \emptyset\}$ we see that $\{R_{\mathbf{i}} : \mathbf{i} \in \mathcal{J}_n\}$ constitutes a partition of $\tilde{\mathcal{E}}_n$. Note that for each $\mathbf{i} \in \mathcal{J}_n$,

$$\lambda(R_{\mathbf{i}}) \leq r_n^d.$$

We claim that for all large n

$$\text{Card}\,(\mathcal{J}_n) \leq C\sqrt{\varepsilon_n} r_n^{-d}. \tag{3.14}$$

To see this, we use the fact that, according to (4.3), there exists $\bar{\rho} > 0$ such that for all large n, $\tilde{\mathcal{E}}_n \subset \mathcal{V}\left(\partial S_f, \bar{\rho}\sqrt{\varepsilon_n}\right)$. Thus, since $r_n/\sqrt{\varepsilon_n} \to 0$ by (3.2),

$$\bigcup_{\mathbf{i} \in \mathcal{J}_n} A_{\mathbf{i}} \subset \mathcal{V}\left(\partial S_f, (\bar{\rho}+2)\sqrt{\varepsilon_n}\right)$$

and, consequently,

$$r_n^d\,\text{Card}\,(\mathcal{J}_n) \leq \lambda\Big(\mathcal{V}\left(\partial S_f, (\bar{\rho}+2)\sqrt{\varepsilon_n}\right)\Big) \leq C\sqrt{\varepsilon_n}.$$

Keeping in mind the fact that for any disjoint sets $B_1, \ldots, B_k$ in $\mathbb{R}^d$ such that, for $1 \leq i \neq j \leq k$,

$$\inf\{\|x - y\| : x \in B_i, y \in B_j\} > r_n,$$

then

$$\int_{B_i} \Delta_n(x)\mathrm{d}x,\ i = 1, \ldots, k, \text{ are independent},$$

we can easily infer that

$$X_{\mathbf{i},n} = \frac{a_n \displaystyle\int_{R_{\mathbf{i}}} \left(\Delta_n(x) - \mathbb{E}\Delta_n(x)\right)\mathrm{d}x}{\sigma_n}, \quad \mathbf{i} \in \mathcal{J}_n,$$

constitutes a 1-dependent random field on $\mathbb{Z}^d$.

Recalling that $a_n = (n/r_n^d)^{1/4}$ and $\sigma_n^2 \to \sigma_f^2$ as $n \to \infty$ by (3.5) we get, for all $\mathbf{i} \in \mathcal{J}_n$,

$$|X_{\mathbf{i},n}| \leq \frac{a_n}{\sigma_n}\lambda(R_{\mathbf{i}}) \leq C(nr_n^{3d})^{1/4}.$$

Hence, by (3.14),

$$\sum_{\mathbf{i} \in \mathcal{J}_n} \mathbb{E}|X_{\mathbf{i},n}|^{5/2} \leq C\left(\text{Card}\,(\mathcal{J}_n)\right)(nr_n^{3d})^{5/8} \leq C(nr_n^{3d/2})^{1/2}.$$

Clearly this bound when combined with (r.iii) and $d \geq 2$, gives as $n \to \infty$,

$$\sum_{\mathbf{i}\in\mathcal{J}_n} \mathbb{E}|X_{\mathbf{i},n}|^{5/2} \to 0,$$

which by the Shergin Fact 3.1 (with $s = 5/2$) yields

$$S_n(\varepsilon_n) = \sum_{\mathbf{i}\in\mathcal{J}_n} X_{\mathbf{i},n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Thus (3.13) holds.

3.3 Central limit theorem for $L_n(\varepsilon_n)$

Now we shall de-Poissonize the central limit for $\Pi_n(\varepsilon_n)$ to obtain one for $L_n(\varepsilon_n)$. Observe that

$$(S_n(\varepsilon_n)|N_n = n) \stackrel{\mathcal{D}}{=} \frac{a_n\,(L_n(\varepsilon_n) - \mathbb{E}\Pi_n(\varepsilon_n))}{\sigma_n}. \tag{3.15}$$

Our next goal is to apply the following version of a theorem in Beirlant and Mason (1995) (see also Polonik and Mason, 2009) to infer from (3.13) that

$$\frac{a_n\,(L_n(\varepsilon_n) - \mathbb{E}\Pi_n(\varepsilon_n))}{\sigma_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{3.16}$$

Fact 3.2 *Let $N_{1,n}$ and $N_{2,n}$ be independent Poisson random variables with $N_{1,n}$ being Poisson $(n\beta_n)$ and $N_{2,n}$ being Poisson $(n(1-\beta_n))$ where $\beta_n \in (0,1)$. Denote $N_n = N_{1,n} + N_{2,n}$ and set*

$$U_n = \frac{N_{1,n} - n\beta_n}{\sqrt{n}} \quad \text{and} \quad V_n = \frac{N_{2,n} - n(1-\beta_n)}{\sqrt{n}}.$$

Let (S_n) be a sequence of real-valued random variables such that

(i) For each $n \geq 1$, the random vector (S_n, U_n) is independent of V_n.

(ii) For some $\sigma^2 < \infty$, $S_n \xrightarrow{\mathcal{D}} \sigma Z$ as $n \to \infty$.

(iii) $\beta_n \to 0$ as $n \to \infty$.

Then, for all x,

$$\mathbb{P}(S_n \leq x \mid N_n = n) \to \mathbb{P}(\sigma Z \leq x).$$

Let

$$\mathcal{D}_n = \{x \in \mathbb{R}^d : f(x) \leq 2\varepsilon_n\}.$$

We shall apply Fact 3.2 to $S_n(\varepsilon_n)$ with

$$N_{1,n} = \sum_{i=1}^{N_n} \mathbf{1}\{X_i \in \mathcal{D}_n\}, \quad N_{2,n} = \sum_{i=1}^{N_n} \mathbf{1}\{X_i \notin \mathcal{D}_n\}$$

and $\beta_n = \mathbb{P}(X \in \mathcal{D}_n)$. Let

$$M = \sup_{x \in \mathbb{R}^d} \sum_{i=1}^{d} \left| \frac{\partial f(x)}{\partial x_i} \right|.$$

We see that for all large enough n, whenever $x \in \mathcal{E}_n$ and $y \in \mathcal{B}(x, r_n)$, by the mean value theorem,

$$f(y) \leq f(x) + Mr_n \leq \varepsilon_n \left(1 + \frac{Mr_n}{\varepsilon_n}\right).$$

This combined with (3.2) implies for all large n

$$\left(\bigcup_{x \in \mathcal{E}_n} \mathcal{B}(x, r_n) \right) \cap \mathcal{D}_n^c = \varnothing.$$

Therefore for all large enough n, the random variables $S_n(\varepsilon_n)$ and $N_{2,n}$ are independent. Thus by (3.15) and $\beta_n \to 0$, we can apply Fact 3.2 to conclude that (3.16) holds.

Next we proceed just as in Mason and Polonik (2009) to apply a moment bound given in Lemma 2.1 of Giné, Mason, and Zaitsev (2003) to show that

$$\mathbb{E}\big(a_n\left(L_n(\varepsilon_n) - \mathbb{E}\Pi_n(\varepsilon_n)\right)\big)^2 \leq 2\sigma_n^2.$$

Therefore, since by (3.5),

$$\sigma_n^2 \to \sigma_f^2 < \infty,$$

the sequence $(a_n(L_n(\varepsilon_n) - \mathbb{E}\Pi_n(\varepsilon_n)))$ is uniformly integrable. Hence we get using (3.16) that

$$a_n\left(\mathbb{E}L_n(\varepsilon_n) - \mathbb{E}\Pi_n(\varepsilon_n)\right) \to 0.$$

Thus, still by (3.16),

$$\frac{a_n\left(L_n(\varepsilon_n) - \mathbb{E}L_n(\varepsilon_n)\right)}{\sigma_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

This in turn implies (3.3).

3.4 Completion of the proof of Theorem 2.1

It remains to verify (3.4). Observe that

$$\overline{L}_n(\varepsilon_n) = \int_{\mathcal{E}_n^c} \Big| \mathbf{1}\{f_n(x) > 0\} - \mathbf{1}\{f(x) > 0\} \Big| \mathrm{d}x = \int_{\mathcal{E}_n^c} \mathbf{1}\{f_n(x) = 0\} \mathrm{d}x.$$

To begin with note that for all $x \in \mathcal{E}_n^c$,

$$\mathbb{P}(f_n(x) = 0) = (1 - \varphi_n(x))^n \leq \exp\left(-n\varphi_n(x)\right),$$

where we recall that

$$\varphi_n(x) = \mathbb{P}\left(X \in \mathcal{B}(x, r_n)\right).$$

Since f is of class $\mathcal{C}^1$, we have, for all $t \in \mathcal{B}(x, r_n)$,

$$|f(t) - f(x)| \leq \kappa_1 r_n,$$

where $\kappa_1 > 0$ is independent of x. Therefore, using the properties $f(x) \geq \varepsilon_n$ and $r_n/\varepsilon_n \to 0$ by (3.2), we obtain

$$\begin{aligned}
\varphi_n(x) = \mathbb{P}\left(X \in \mathcal{B}(x, r_n)\right) &= f(x)\omega_d r_n^d + \int_{\mathcal{B}(x,r_n)} \left(f(t) - f(x)\right) \mathrm{d}t \\
&\geq \varepsilon_n r_n^d \left(\omega_d - \kappa_1 \frac{r_n}{\varepsilon_n}\right) \\
&\geq \kappa_2 \varepsilon_n r_n^d,
\end{aligned}$$

where $\kappa_2 > 0$ is independent of x. Thus,

$$\mathbb{P}(f_n(x) = 0) \leq \exp(-\kappa_2 \varepsilon_n n r_n^d).$$

Consequently,

$$a_n \, \mathbb{E}\overline{L}_n(\varepsilon_n) \leq \left(\frac{n}{r_n^d}\right)^{1/4} \exp(-\kappa_2 \varepsilon_n n r_n^d).$$

By (r.ii), for all n large enough, $r_n^{-d} \leq n$. Hence,

$$\frac{n}{r_n^d} \leq n^2.$$

Besides, provided n is large enough, we get by (r.ii) that $nr_n^d \geq (\ln n/\kappa_2)^{4/3}$. Consequently,

$$\varepsilon_n n r_n^d = (nr_n^d)^{3/4} \geq \frac{1}{\kappa_2} \ln n.$$

This gives the bound holding for all large n,

$$\left(\frac{n}{r_n^d}\right)^{1/4} \exp\left(-\kappa_2 \varepsilon_n n r_n^d\right) \leq \sqrt{n} \exp\left(-\ln n\right) = n^{-1/2},$$

which goes to 0 as $n \to \infty$. This implies that both $a_n \mathbb{E}\overline{L}_n(\varepsilon_n) \to 0$ and $a_n \overline{L}_n(\varepsilon_n) \overset{\mathbb{P}}{\to} 0$, and thus establishes (3.4). The proof of Theorem 2.1 now follows from (3.3) and (3.4).

4 Appendix: Properties of $\{x : 0 < f(x) \leq \varepsilon\}$

Under Assumption Sets 1 and 2, we know that there exists a tubular neighborhood $\mathcal{V}(\partial S_f, \rho)$ of ∂S_f of radius ρ such that first,

$$0 < \inf_{p \in \partial S_f} \inf_{0 \leq u \leq \rho} D_{e_p}^2 f(p + u e_p) \leq \sup_{p \in \partial S_f} \sup_{0 \leq u \leq \rho} D_{e_p}^2 f(p + u e_p) < \infty, \tag{4.1}$$

and second,

$$\inf\{f(x) \,:\, x \in S_f \backslash \mathcal{V}(\partial S_f, \rho)\} = \sup\{f(x) \,:\, x \in \mathcal{V}(\partial S_f, \rho)\} := \varepsilon_0 > 0.$$

Consequently, for all $0 < \varepsilon < \varepsilon_0$, we have

$$\{x \in \mathbb{R}^d \,:\, 0 < f(x) \leq \varepsilon\} \subset \mathcal{V}(\partial S_f, \rho).$$

Moreover, (4.1), together with the fact that $f = 0$ on ∂S_f, entails that for all $p \in \partial S_f$, the maps $u \mapsto f(p + u e_p)$ are strictly convex and strictly increasing on $[0, \rho]$. Therefore, for all $0 < \varepsilon < \varepsilon_0$, and for all $p \in \partial S_f$ there exists a unique real number $\kappa_p(\varepsilon)$ such that

$$f(p + \kappa_p(\varepsilon) e_p) = \varepsilon.$$

Note that we also have the relation

$$\bigcup_{p \in \partial S_f} \{p + u e_p : 0 \leq u \leq \kappa_p(\varepsilon)\} = \{x : 0 < f(x) \leq \varepsilon\} \cup \partial S_f, \tag{4.2}$$

for all $0 < \varepsilon < \varepsilon_0$.

Since $D_{e_p} f(p) = 0$ for all $p \in \partial S_f$ by assumption, by using a second order expansion of f at p in combination with (4.1), we have

$$\sup_{p \in \partial S_f} \kappa_p(\varepsilon) \leq \left[\frac{1}{2} \inf_{p \in \partial S_f} \inf_{0 \leq u \leq \rho} D_{e_p}^2 f(p + u e_p)\right]^{-\frac{1}{2}} \sqrt{\varepsilon}. \tag{4.3}$$

The same argument gives

$$\inf_{p \in \partial S_f} \kappa_p(\varepsilon) \geq \left[\frac{1}{2} \sup_{p \in \partial S_f} \sup_{0 \leq u \leq \rho} D_{\varepsilon_p}^2 f(p + u e_p)\right]^{-1/2} \sqrt{\varepsilon}. \tag{4.4}$$

Acknowledgments. We thank two referees for valuable comments and insightful suggestions.

References

[1] Baíllo, A. and Cuevas, A. (2006). Image estimators based on marked bins, *Statistics*, **Vol. 40**, pp. 277–288.

[2] Baíllo, A., Cuevas, A., and Justel, A. (2000). Set estimation and nonparametric detection, *Canadian Journal of Statistics*, **Vol. 28**, pp. 765-782.

[3] Beirlant, J., Györfi, L., and Lugosi, G. (1994). On the asymptotic normality of the L_1- and L_2-errors in histogram density estimation, *Canadian Journal of Statistics*, **Vol. 22**, pp. 309-318.

[4] Beirlant, J. and Mason, D.M. (1995). On the asymptotic normality of L_p-norms of empirical functionals, *Mathematical Methods of Statistics*, **Vol. 4**, pp. 1-19.

[5] Biau, G., Cadre, B., and Pelletier, B. (2008). Exact rates in density support estimation, *Journal of Multivariate Analysis*, **Vol. 99**, pp. 2185-2207.

[6] Bredon, G.E. (1993). *Topology and Geometry*, Volume 139 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

[7] Cuevas, A., Fraiman, R., and Rodríguez-Casal, A. (2007). Nonparametric approach to the estimation of lengths and surface areas, *The Annals of Statistics*, **Vol. 35**, pp. 1031–1051.

[8] Cuevas, A. and Rodríguez-Casal, A. (2004). On boundary estimation, *Advances in Applied Probability*, **Vol. 36**, pp. 340–354.

[9] Devroye, L. and Wise, G. (1980). Detection of abnormal behavior via nonparametric estimation of the support, *SIAM Journal on Applied Mathematics*, **Vol. 38**, pp. 480-488.

[10] Gallot, S., Hulin, D. and Lafontaine, J. (2004). *Riemannian Geometry*, 3rd Ed., Springer-Verlag, New York.

[11] Giné, E., Mason, D.M., and Zaitsev, A. (2003). The L_1-norm density estimator process, *The Annals of Probability*, **Vol. 31**, pp. 719-768.

[12] Gray, A. (1990). *Tubes*, Addison-Wesley, Redwood City, California.

[13] Hall, P. (1984). Random, nonuniform distribution of line segments on a circle, *Stochastic Processes and their Applications*, **Vol. 18**, pp. 239-261.

[14] Hall, P. (1985). Three limit theorems for vacancy in multivariate coverage problems, *Journal of Multivariate Analysis*, **Vol. 16**, pp. 211-236.

[15] Hall, P. (1988). *Introduction to the Theory of Coverage Processes*, John Wiley & Sons, New York.

[16] Härdle, W., Park, B.U., and Tsybakov, A.B. (1995). Estimation of non-sharp support boundaries, *Journal of Multivariate Analysis*, **Vol. 55**, pp. 205–218.

[17] Mason, D.M. and Polonik, W. (2009). Asymptotic normality of plug-in level set estimates, *The Annals of Applied Probability*, **Vol. 19**, pp. 1108–1142 .

[18] Molchanov, I.S. (1998). A limit theorem for solutions of inequalities, *Scandinavian Journal of Statistics*, **Vol. 25**, pp. 235–242.

[19] Shergin, V.V. (1990). The central limit theorem for finitely dependent random variables. In: *Proc. 5th Vilnius Conf. Probability and Mathematical Statistics, Vol. II*, Grigelionis, B., Prohorov, Y.V., Sazonov, V.V., and Statulevicius, V. (Eds.), pp. 424-431.