SUPPLEMENTAL APPENDIX FOR "NONPARAMETRIC ESTIMATION IN RANDOM COEFFICIENTS BINARY CHOICE MODELS

ERIC GAUTIER AND YUICHI KITAMURA

A.1.1. **A Toy Model.** As noted in the main text, the key insight for our estimation procedure lies in the fact the estimation of f_β in (1.3) is mathematically equivalent to a statistical deconvolution problem. To see this, it is useful to first consider the case with $d = 2$. We parameterize the vectors $b = (b_1, b_2)'$ and $x = (x_1, x_2)'$ on $\mathbb{S}^1$ by their angles $\phi = \arccos(b_1)$ and $\theta = \arccos(x_1)$ in $[0, 2\pi)$. As is often the case when Fourier series techniques are used, we consider spaces of complex valued functions. Let $\mathrm{L}^p(\mathbb{S}^1)$ denote the Banach space of Lebesgue p-integrable functions and its norm by $\|\cdot\|_p$. In the case of $\mathrm{L}^2(\mathbb{S}^1)$, the norm is derived from the hermitian product $\int_0^{2\pi} f(\theta)\overline{g(\theta)}d\theta$. Let R_θ and f_ϕ denote the extension R of r according to (3.2) and f_β after the reparameterization. Our task is then to obtain f_ϕ from the knowledge of R_θ. Rewrite (1.3) using these definitions, then divide both sides by π, to get:

$$\text{(A.1)} \qquad \frac{R_\theta}{\pi}(\theta) = \frac{\mathcal{H}(f_\beta)}{\pi}(\theta) = \int_0^{2\pi} \left(\frac{1}{\pi}\mathbb{I}\left\{|\theta - \phi| < \pi/2\right\} \right) f_\phi(\phi)d\phi.$$

If we further define $f_\theta := R_\theta/\pi$ and $f_\eta(\eta) := \frac{1}{\pi}\mathbb{I}\{|\eta| < \pi/2\}$, then using the standard notation for convolution, (A.1) can be written as $f_\theta = f_\eta * f_\phi$. It is now obvious that the estimation of f_ϕ (thus f_β) is linked to the following statistical deconvolution problem: unobservable random variables ϕ and η with densities f_ϕ and f_η are related to an observable random variable θ according to $\theta = \eta + \phi$, and one wishes to recover f_ϕ from f_θ, the density of θ, when f_η is known (and it is Uniform$[-\pi/2, \pi/2]$ in this case).[1]

The problem of deconvolution on the unit circle can be conveniently solved using Fourier series. The set of functions $\left(\exp(-int)/\sqrt{2\pi}\right)_{n\in\mathbb{Z}}$ is the orthonormal basis of $\mathrm{L}^2(\mathbb{S}^1)$ used to define Fourier series. This system is also complete in $\mathrm{L}^1(\mathbb{S}^1)$. Reparameterize a function $f \in \mathrm{L}^1(\mathbb{S}^1)$ it using angles as above, and denote it by f_t. Denoting the Fourier coefficients of $f \in \mathrm{L}^1(\mathbb{S}^1)$ by

[1]It is also useful to note that the inversion of $\mathcal{H}$ is closely related to differentiation. Differentiating the right hand-side of expression (A.1) with respect to θ identifies $f_\phi(\theta + \pi/2) - f_\phi(\theta - \pi/2)$ where f_ϕ is defined on the line by periodicity. If f_ϕ is supported on a semicircle, with an assumption that is elaborated further in Section 3, f_ϕ (which is positive) is identified. Thus if the model is identified the inverse of $\mathcal{H}$ is a differential operator and as such unbounded.

$c_n(f_t) = \int_0^{2\pi} f_t(t)\exp(-int)dt/(2\pi)$,

$$f_t(\theta) = \sum_{n\in\mathbb{Z}} c_n(f_t)\exp(int) \tag{A.2}$$

holds in the $\mathrm{L}^1(\mathbb{S}^1)$ sense. Recall also that for f and g in $\mathrm{L}^1(\mathbb{S}^1)$, after the same reparameterization,

$$c_n(f_t * g_t) = 2\pi c_n(f_t)c_n(g_t). \tag{A.3}$$

Using equation (A.3) we obtain the following proposition.

Proposition A.1. *$c_0(R_\theta) = \pi c_0(f_\phi)$ and for $n \in \mathbb{Z}\setminus\{0\}$, $c_n(R_\theta) = c_n(f_\phi)\,2\sin(n\pi/2)/n$.*

As in classical deconvolution problems on the real line, our aim is to obtain f_t (thus f_β) using equation (A.2) and Proposition A.1. Proposition A.1 shows that $c_{2p}(R_\theta) = 0$ holds for all non-zero p's, regardless of the values of $c_{2p}(f_\phi), p \in \mathbb{Z}\backslash\{0\}$. Thus from $r(x) = R_\theta(\theta)$ one can only recover the Fourier coefficients $c_n(f_\phi)$ for $n = 0$ (which is easily seen to be $1/2\pi$, by integrating both sides of (A.1) and noting that f_β is a probability density function) and $n = 2p+1, p \in \mathbb{Z}$. The same phenomenon occurs in higher dimensions, as explained in Section A.1.3.

Remark A.1. The vector spaces $H^{2p+1,2} = \mathrm{span}\left\{\exp(i(2p+1)t)/\sqrt{2\pi}, \exp(-i(2p+1)t)/\sqrt{2\pi}\right\}, p \in \mathbb{N}$ are eigenspaces of the compact self-adjoint operator $\mathcal{H}$ on $\mathrm{L}^2(\mathbb{S}^1)$. These eigenspaces are associated with the eigenvalues $\frac{2(-1)^p}{2p+1}$. Also, $\bigoplus_{p\in\mathbb{N}\backslash\{0\}} H^{2p,2}$ is the null space $\ker \mathcal{H}$.

A.1.2. **The Gegenbauer polynomials.** We summarize some results on the Gegenbauer polynomials, which are used in various parts of the paper. These can be found in Erdélyi et al. (1953) and Groemer (1996). When $\nu = 0$ and $d = 2$, it is related to the Chebychev polynomials of the first kind, as

$$\forall n \in \mathbb{N}\setminus\{0\},\ C_n^0(t) = \frac{2}{n}T_n(t)$$

and

$$C_0^0(t) = T_0(t) = 1$$

hold for

$$T_n(t) = \cos\left(n\arccos(t)\right), n \in \mathbb{N}.$$

When $\nu = 1$ and $d = 4$, $C_n^1(t)$ coincides with the Chebychev polynomial of the second kind $U_n(t)$, which is given by

$$U_n(t) = \frac{\sin[(n+1)\arccos(t)]}{\sin[\arccos(t)]}, n \in \mathbb{N}.$$

The Gegenbauer polynomials are orthogonal with respect to the weight function $(1-t^2)^{\nu-1/2}dt$ on $[-1,1]$. Note that $C_0^\nu(t)=1$ and $C_1^\nu(t)=2\nu t$ for $\nu\neq 0$ while $C_1^0(t)=2t$. Moreover, the following recursion relation holds

$$(n+2)C_{n+2}^\nu(t)=2(\nu+n+1)tC_{n+1}^\nu(t)-(2\nu+n)C_n^\nu(t). \tag{A.4}$$

Implementation of our estimator requires evaluation of the Gegenbauer polynomials for a series of successive values of n. The recursion relation (A.4) is therefore a powerful tool. The Gegenbauer polynomials are related to each other through differentiation, that is, they satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}C_n^\nu(t)=2\nu C_{n-1}^{\nu+1}(t) \tag{A.5}$$

for $\nu>0$ and

$$\frac{\mathrm{d}}{\mathrm{d}t}C_n^0(t)=2C_{n-1}^1(t). \tag{A.6}$$

For $\nu\neq 0$ the Rodrigues formula states that

$$C_n^\nu(t)=(-2)^{-n}(1-t^2)^{-\nu+1/2}\frac{(2\nu)_n}{(\nu+1/2)_n n!}\frac{\mathrm{d}^n}{\mathrm{d}t^n}(1-t^2)^{n+\nu-1/2}. \tag{A.7}$$

The following results are also used in the paper:

$$\sup_{t\in[-1,1]}\left|\frac{C_n^\nu(t)}{C_n^\nu(1)}\right|\leq 1, \tag{A.8}$$

$$\forall\ \nu>0,\ \forall n\in\mathbb{N},\ C_n^\nu(1)=\begin{pmatrix} n+2\nu-1 \\ n \end{pmatrix} \tag{A.9}$$

$$C_0^0(1)=1 \text{ and } \forall n\in\mathbb{N}\setminus\{0\},\ C_n^0(1)=\frac{2}{n}, \tag{A.10}$$

$$C_n^\nu(-t)=(-1)^n C_n^\nu(t). \tag{A.11}$$

These orthogonal polynomials are normalized such that

$$\|C_n^{\nu(d)}\left(x'\cdot\right)\|_2^2=|\mathbb{S}^{d-2}|\int_{-1}^1(C_n^{\nu(d)}(t))^2(1-t^2)^{(d-3)/2}dt=\frac{|\mathbb{S}^{d-1}|(C_n^{\nu(d)}(1))^2}{h(n,d)}. \tag{A.12}$$

A.1.3. **Tools for Higher Dimensional Spheres.** Let us introduce some concepts used for the treatment of the general case $d \geq 2$. We consider functions defined on the sphere $\mathbb{S}^{d-1}$, which is a $d-1$ dimensional smooth submanifold in $\mathbb{R}^d$. The canonical measure on $\mathbb{S}^{d-1}$ (or the spherical measure) is denoted by σ. It is a uniform measure on $\mathbb{S}^{d-1}$ satisfying $\int_{\mathbb{S}^{d-1}} d\sigma = |\mathbb{S}^{d-1}|$, where $|\mathbb{S}^{d-1}|$ signifies the surface area of the unit sphere.

Recall that the basis functions $\exp(\pm int)/\sqrt{2\pi}$ are eigenfunctions of $-\frac{\mathrm{d}}{\mathrm{d}t^2}$ associated with eigenvalue n^2. In a similar way, the Laplacian on the sphere $\mathbb{S}^{d-1}, d \geq 2$, denoted by Δ^S, can be used to obtain an orthonormal basis for higher dimensional spheres. It can be defined by the formula

$$\Delta^S f = (\Delta \breve{f})\check{} \tag{A.13}$$

where Δ is the Laplacian in $\mathbb{R}^d$, $\breve{f}$ the radial extension of f, that is $\breve{f}(x) = f(x/\|x\|)$, and $\check{f}$ the restriction of f to $\mathbb{S}^{d-1}$. Likewise the gradient on the sphere is given by:

$$\nabla^S f = (\nabla \breve{f})\check{} \tag{A.14}$$

where ∇ is the gradient in $\mathbb{R}^d$.

Definition A.1. A surface harmonic of degree n is the restriction of a homogeneous harmonic polynomial (a homogeneous polynomial p whose Laplacian Δp is zero) of degree n in $\mathbb{R}^d$ to $\mathbb{S}^{d-1}$.

The reader is referred to Müller (1966) and Groemer (1996) for clear and detailed expositions on these concepts and important results concerning spherical harmonics used in this paper. Erdélyi et al. (1953, vol. 2, chapter 9) provide detailed accounts focusing on special functions. Here are some useful results:

Lemma A.1. *The following properties hold:*

(i) $-\Delta^S$ *is a positive self-adjoint unbounded operator on* $\mathrm{L}^2(\mathbb{S}^{d-1})$, *thus it has orthogonal eigenspaces and a basis of eigenfunctions;*

(ii) Surface harmonics of degree n *are eigenfunctions of* $-\Delta^S$ *for the eigenvalue* $\zeta_{n,d} := n(n+d-2)$;

(iii) The dimension of the vector space $H^{n,d}$ *of surface harmonics of degree* n *is*

$$h(n,d) := \frac{(2n+d-2)(n+d-2)!}{n!(d-2)!(n+d-2)}; \tag{A.15}$$

(iv) A system formed of orthonormal bases $(Y_{n,l})_{l=1}^{h(n,d)}$ *of* $H^{n,d}$ *for each degree* $n = 0, \dots, \infty$ *is complete in* $\mathrm{L}^1(\mathbb{S}^{d-1})$, *that is, for every* $f \in \mathrm{L}^1(\mathbb{S}^{d-1})$ *the following equality holds in the* $\mathrm{L}^1(\mathbb{S}^{d-1})$

sense:

$$f = \sum_{n=0}^{\infty} \sum_{l=1}^{h(n,d)} (f, Y_{n,l})_{\mathrm{L}^2(\mathbb{S}^{d-1})} Y_{n,l}.$$

Thus $h(n,d)$ is the multiplicity of the eigenvalue $\zeta_{n,d}$, and $H^{n,d}$ is the corresponding eigenspace. Lemma A.1 (i), (ii) and (iv) give the decomposition

$$\mathrm{L}^2(\mathbb{S}^{d-1}) = \bigoplus_{n\in\mathbb{N}} H^{n,d}.$$

The space of surface harmonics of degree 0 is the one dimensional space spanned by 1. A series expansion on an orthonormal basis of surface harmonics is called a Fourier series when $d=2$, a Laplace series when $d=3$ and in the general case a Fourier-Laplace series.

Orthonormal bases of surface harmonics usually involve parametrization by angles, such as the spherical coordinates when $d=3$ as used by Healy and Kim (1996) or hyperspherical coordinates for $d>3$. Instead, here we work with the decomposition of a function on the spaces $H^{n,d}$ as presented in the next definition so that we avoid specific expressions of basis functions.

Definition A.2. The condensed harmonic expansion of a function f in $\mathrm{L}^1(\mathbb{S}^{d-1})$ is the series $\sum_{n=0}^{\infty} Q_{n,d} f$, where $Q_{n,d}$ is the projector from $\mathrm{L}^2(\mathbb{S}^{d-1})$ to $H^{n,d}$.

This leads to a simple method both in terms of theoretical developments and practical implementations. The projector $Q_{n,d}$ can be expressed as an integral operator with kernel

$$q_{n,d}(x,y) = \sum_{l=1}^{h(n,d)} \overline{Y_{n,l}(x)} Y_{n,l}(y), \tag{A.16}$$

where $(Y_{n,l})_{l=1}^{h(n,d)}$ is any orthonormal basis of $H^{n,d}$. The kernel has a simple expression given by the addition formula:

Theorem A.1 (Addition Formula)**.** *For every x and $y \in \mathbb{S}^{d-1}$, we have*

$$q_{n,d}(x,y) = {}^{\flat}q_{n,d}(x'y), \quad {}^{\flat}q_{n,d}(t) := \frac{h(n,d) C_n^{\nu(d)}(t)}{|\mathbb{S}^{d-1}| C_n^{\nu(d)}(1)} \tag{A.17}$$

where C_n^{ν} are Gegenbauer a and $\nu(d) = (d-2)/2$.

The Sobolev spaces are defined in the Fourier-Laplace domain through the fractional Laplacian defined on a certain subset of $\mathrm{L}^p(\mathbb{S}^{d-1})$ as

$$\left(-\Delta^S\right)^{s/2} f := \sum_{n=0}^{\infty} \zeta_{n,d}^{s/2} Q_{n,d} f. \tag{A.18}$$

For the case where $p = 2$, in stead of the definition of the norm $\|\cdot\|_{p,s}$ given in Section 3 it is also possible to use an equivalent norm, the square of which is equal to

$$\sum_{n=0}^{\infty}(1+\zeta_{n,d})^s\,\|Q_{n,d}f\|_2^2.$$

The following integration by parts holds for functions f in $\mathrm{H}^1(\mathbb{S}^{d-1})$

$$-\int_{\mathbb{S}^{d-1}} f(x)\Delta^S f(x)d\sigma(x) = \int_{\mathbb{S}^{d-1}} \nabla_x^S f'\nabla_x^S f d\sigma(x) \tag{A.19}$$

and as a consequence for the second definition of the norm of $\mathrm{H}^1(\mathbb{S}^{d-1})$ we have

$$\|f\|_{2,1}^2 = \|f\|_2^2 + \|\nabla^S f\|_2^2.$$

In Section A.1.1 we observed the close relationship between the random coefficients binary choice model and convolution for $d = 2$. This connection remains valid in higher dimensions. Suppose a function $f(x, y)$ defined on $\mathbb{S}^{d-1}\otimes\mathbb{S}^{d-1}$ depends on x and y only through the spherical distance $d(x, y) = \arccos(x'y)$ (that is, f is a zonal function). Consider the following integral:

$$h(x) = \int_{\mathbb{S}^{d-1}} f(x, y)g(y)d\sigma(y) := f * g(x),$$

then the function h is a convolution on the sphere. We now see that the choice probability function $r(x) = \mathcal{H}(f_\beta)(x) = \int_{\mathbb{S}^{d-1}} \mathbb{I}\{x'b \geq 0\}f_\beta(b)d\sigma(b)$ is a special case of h and therefore can also be regarded as convolution. Obtaining f_β from r (or, inverting $\mathcal{H}$) is therefore a deconvolution problem.

In what follows we often write $f(x,\star)$ when a function f on $\mathbb{S}^{d-1}\otimes\mathbb{S}^{d-1}$ is regarded as a function of $\star$. Also, the notation $\|f(x,\star)\|_p$ is used for the L^p norm of $f(x,\star)$, that is, $\|f(x,\star)\|_p = \int_{\mathbb{S}^{d-1}} |f(x,y)|^p d\sigma(y)$. Note that if f is a zonal function as in the above definition of spherical convolution, its L^p norm $\|f(x,\star)\|_p$ does not depend on x. The following Young inequalities for convolution on the sphere (see, for example, Kamzolov, 1983) are useful:

Proposition A.2 (Young inequalities)**.** *Suppose $f(x,\star)$ and g belong to $\mathrm{L}^r(\mathbb{S}^{d-1})$ and $\mathrm{L}^p(\mathbb{S}^{d-1})$, respectively. Then $h(x) = f * g(x)$ is well-defined in $\mathrm{L}^q(\mathbb{S}^{d-1})$ and*

$$\|h\|_q \leq \|f\|_r\|g\|_p,$$

where $1 \leq p, q, r \leq \infty$ and $\frac{1}{q} = \frac{1}{p} + \frac{1}{r} - 1$.

Let P_T denote the projection operator onto $\bigoplus_{n=0}^{T} H^{n,d}$, i.e.

$$P_T f(x) = \sum_{n=0}^{T} Q_{n,d}f(x) = \int_{\mathbb{S}^{d-1}} D_T(x, y)f(y)d\sigma(y) \tag{A.20}$$

where

$$D_T(x,y)=\sum_{n=0}^{T} q_{n,d}(x,y).$$

The kernel D_T extends the classical Dirichlet kernel on the circle to the sphere $\mathbb{S}^{d-1}$. The sum over T in the definition of D_T also has the simple closed form in terms of derivatives of Gegenbauer polynomials; see Equation (52) in Müller (1966). The linear form $f \to \int_{\mathbb{S}^{d-1}} D_T(x,y)f(y)d\sigma(y)$ converges to $\int_{\mathbb{S}^{d-1}} f(y)d\delta_x(y)=f(x)$ as T goes to infinity, where δ_x denotes the Dirac measure. The Dirichlet kernel yields the best approximation $P_T f$ of f in $\mathrm{L}^2(\mathbb{S}^{d-1})$ by polynomials that belong to $\bigoplus_{n=0}^{T} H^{n,d}$, but is known to have flaws. For example, D_T does not satisfy

$$\forall f \in \mathrm{L}^1(\mathbb{S}^{d-1}), \lim_{T\to\infty} \|D_T * f - f\|_{\mathrm{L}^1(\mathbb{S}^{d-1})}=0,$$

that is, the sequence $D_T, T=0,1,...$ is not an approximate identity (see, e.g., Devroye and Gyorfi 1985) in $\mathrm{L}^1(\mathbb{S}^{d-1})$. Indeed, the $\mathrm{L}^1(\mathbb{S}^{d-1})$ norm of the kernel is not uniformly bounded; more precisely, we have

$$\|D_T(\cdot,x)\|_1 \asymp T^{(d-2)/2} \tag{A.21}$$

when $d \geq 3$ and

$$\|D_T(\cdot,x)\|_1 \asymp \log T \tag{A.22}$$

when $d=2$ (as noted above, these norms do not depend on the value of $x \in \mathbb{S}^{d-1}$). These bounds can be found in Gronwall (1914) for $d=3$ and Ragozin (1972) and Colzani and Traveglini (1991) for higher dimensions. Also, D_T does not have good approximation properties in $\mathrm{L}^\infty(\mathbb{S}^{d-1})$; in particular, we do not have

$$\forall f \in \mathrm{L}^\infty(\mathbb{S}^{d-1}), \lim_{T\to\infty} \|D_T * f - f\|_{\mathrm{L}^\infty(\mathbb{S}^{d-1})}=0.$$

Near the points of discontinuity of f, $D_T * f$ has oscillations which do not decay to zero as T grows to infinity, known as the Gibbs oscillations. This phenomenon deteriorates as the dimension increases. These problems can be addressed by using kernels that involves extra smoothing instead of the Dirichlet kernel D_T. To this end, define a general class of kernel

$$K_T(x,y)=\sum_{n=0}^{T} \chi(n,T) q_{n,d}(x,y) \tag{A.23}$$

for some sequence $\chi(n,T)$. These are called smoothed projection kernels. Typically the function χ is chosen so that it puts more weight on lower frequencies. In particular we impose the following conditions:

Assumption A.1. *(i)* $\|K_T(x,\star)\|_1$ *is uniformly bounded in* T.

(ii) *There exists constants* C *and* α *such that for all* $x, y, z \in \mathbb{S}^{d-1}$,

$$|K_T(z,x) - K_T(z,y)| \leq C\|x-y\|T^{\alpha},$$

where $\|\cdot\|$ *denotes the Euclidean norm.*

(iii) *For* $p \in [1,\infty]$ *and* $s > 0$, *there exists a constant* C *such that for every* f *in* $\mathrm{W}_p^s(\mathbb{S}^{d-1})$,

$$\left\| f(\cdot) - \int_{\mathbb{S}^{d-1}} K_T(\cdot,y)f(y)d\sigma(y) \right\|_p \leq CT^{-s}\|f\|_{p,s}.$$

(iv) $\chi(\cdot,T)$ *takes values in* $[0,1]$ *and is such that there exists* $c > 0$ *such that for all* $0 \leq n \leq \lfloor T/2 \rfloor$, $\chi(n,T) \geq c$.

The smoothed projection kernel $K_T(x,y)$ depends on x and y only through $d(x,y)$, thus the value of the norm $\|K_T(x,\star)\|_1$ in Assumption (i) does not depend on $x \in \mathbb{S}^{d-1}$. Assumption (i) could be relaxed, but imposing this on K_T allows us to make relatively weak assumptions on the smoothness of the density of the covariates later in this paper. Assumption (ii) is used to establish the L^{∞}-rates of convergence of our estimators. Assumption (iii) provides bounds for approximation errors. Under this condition, $K_T * f$ approximates $f \in \mathrm{L}^p(\mathbb{S}^{d-1})$ with an error of the same order as that of the best n-th degree spherical harmonic approximation of a function $f \in \mathrm{L}^p(\mathbb{S}^{d-1})$ in $\mathrm{W}_p^s(\mathbb{S}^{d-1})$ (see e.g. Kamzolov 1983 and Ditzian 1998). This is useful in our treatment of the bias terms in our estimators. As concrete examples, the following two choices for the weight function χ in (A.23) satisfy Assumption A.1, as shown in the appendix. The first and the second choices of χ correspond to the *Riesz kernel* and the *delayed means kernel*, respectively.

Proposition A.3. *In the definition of the smoothed kernel* (A.23), *let*

$$\chi(n,T) = \left(1 - \left(\frac{\zeta_{n,d}}{\zeta_{T,d}+1}\right)^{s/2}\right)^{l},$$

where l *is an integer satisfying* $l > (d-2)/2$, *or*

$$\chi(n,T) = \psi(n/T)$$

where $\psi: [0,\infty) \to [0,\infty)$ *is infinitely differentiable, nonincreasing, such that* $\psi(x) = 1$ *if* $x \in [0,1]$, $0 \leq \Psi(x) \leq 1$ *if* $x \in [1,2]$, $\psi(x) = 0$ *if* $x \geq 2$. *Then* K_T *satisfies Assumption A.1.*

The delayed means kernel has the nice property that it does not require prior knowledge of the regularity s in Assumption A.1. The Dirichlet kernel satisfies (ii), (iii) (for $p = 2$) and (iv) of

Assumption A.1. Like the delayed means kernel, it achieves the optimal rate of approximation without the prior knowledge of s.

Proof of Proposition A.3. First consider the Riesz kernel. (i) follows from (2.4) in Ditzian (1998) and by the fact that Cesàro kernels C_h^l are uniformly bounded in $\mathrm{L}^1(\mathbb{S}^{d-1})$ for $l > \frac{d-2}{2}$ (see, e.g. Bonami and Clerc 1973, p. 225). To show (iii) we use Theorem 4.1 in Ditzian (1998), by letting $P(D) = \Delta^S$, $\lambda = \zeta_{T,d} + 1 = T(T + d - 2) + 1$, $\alpha = s/2$ and $m = 1$. Then it implies an approximation error upper bound $CK_{s/2}(f, \Delta^S, (\zeta_{T,d}+1)^{-\frac{s}{2}})$, which, in turn, is bounded by $CT^{-s}\|(-\Delta^S)^{s/2}f\|_p$ (see equations (4.2) and (4.1) therein). By the definition of the norm of the Sobolev space $W_p^s(\mathbb{S}^{d-1})$ (see (3.6)) the result follows. Concerning the delayed means, (i) follows from Theorem 2.2 and Proposition 2.5 of Narcowich et al. (2006). (ii) corresponds to Lemma 2.6 in Narcowich et al. (2006). To see (iii), use Lemma 2.4 (c) in Narcowich et al. (2006) to obtain an upper bound $C \inf_{g\in\bigoplus_{n=0}^{T/2} H^{n,d}} \|f - g\|_p$. Let $\lambda = \zeta_{T/2,d} + 1 = \frac{T}{2}(\frac{T}{2} + d - 2) + 1$, $\alpha = s/2, m = 1, P(D) = \Delta^S$ in Ditzian's (1998) Theorem 6.1, which gives an upper bound on the best spherical harmonic approximation in $\mathrm{L}^p(\mathbb{S}^{d-1})$ to functions in $\mathrm{W}_p^s(\mathbb{S}^{d-1})$ (see also Kamzolov, 1983), then apply equation (4.1) in Ditzian (1998) again to obtain the desired result. □

If the function f is in $\mathrm{L}^2(\mathbb{S}^{d-1})$ then Equations (A.17) and (A.11) imply that $Q_{2p,d}f(x) = Q_{2p,d}f(-x)$ and $Q_{2p+1,d}f(x) = -Q_{2p+1,d}f(-x)$ for $p \in \mathbb{N}$. Consequently, the odd order terms in the condensed harmonic expansions of f, f^+ and f^- satisfy $Q_{2p+1}f^- = Q_{2p+1}f$ and $Q_{2p+1}f^+ = 0$. Likewise, for the even order terms in the condensed harmonic expansions of these functions $Q_{2p}f^+ = Q_{2p}f$ and $Q_{2p}f^- = 0$ hold. We conclude that the sum of the odd order terms in the condensed harmonic expansion corresponds to f^- and that of the even order terms to f^+. As anticipated from the analysis of the $d = 2$ case, the operator $\mathcal{H}$ reduces the even part of f_β to a constant $\frac{1}{2}$, therefore Fourier-Laplace series expansions for f_β derived later involve only odd order terms.

We now provide a formula that is used to obtain our estimator for f_β. If a non-negative function f has its support included in some hemisphere of $\mathbb{S}^{d-1}$ then

$$f(x) = 2f^-(x)\mathbb{I}\left\{f^-(x) > 0\right\}. \tag{A.24}$$

Denote the support of f by supp$f$ and let $-\mathrm{supp}f = \{x| - x \in \mathrm{supp}f\}$, then this formula follows from the fact that $f^-(x) = f^+(x) \geq 0$ on supp$f$ while $f^-(x) = -f^+(x) \leq 0$ on $-$supp$f$ and both f^- and f^+ are 0 on $\mathbb{S}^{d-1} \setminus (\mathrm{supp}f \bigcup -\mathrm{supp}f)$.

Remark A.2. If f is a probability density function, the coefficient of degree 0 in the expansion of f on surface harmonics is $1/|\mathbb{S}^{d-1}|$. Conversely, any harmonic polynomial or series such that its degree 0 coefficient is $1/|\mathbb{S}^{d-1}|$ integrates to one.

The next theorem shows that Fourier-Laplace series on the sphere is a natural tool for the study of the operator $\mathcal{H}$.

Theorem A.2 (Funk-Hecke Theorem)**.** *If g belongs to $H^{n,d}$ for some n, and a function F on $(-1,1)$ satisfies*

$$\int_{-1}^{1} |F(t)|^2(1-t^2)^{(d-3)/2}dt < \infty,$$

then

$$\int_{\mathbb{S}^{d-1}} F(x'y)g(y)d\sigma(y) = \lambda_n(F)g(x) \tag{A.25}$$

where

$$\lambda_n(F) = |\mathbb{S}^{d-2}|C_n^{\nu(d)}(1)^{-1}\int_{-1}^{1} F(t)C_n^{\nu(d)}(t)(1-t^2)^{\frac{d-3}{2}}dt.$$

In other words, the kernel operator defined by

$$f \in \mathrm{L}^2(\mathbb{S}^{d-1}) \mapsto \left(x \mapsto \int_{\mathbb{S}^{d-1}} F(x'y)f(y)d\sigma(y)\right) \in \mathrm{L}^2(\mathbb{S}^{d-1})$$

is, in the subspace $H^{n,d}$, equivalent to the multiplication by $\lambda_n(F)$. Thus a basis of surface harmonics diagonalizes an integral operator if its kernel is a function of the scalar product $x'y$.

Remark A.3. Healy and Kim (1996) use Fourier-Laplace expansions to analyze a deconvolution problem on $\mathbb{S}^2$. As we shall see below, the Addition Formula along with condensed harmonic expansions provide a general treatment that works for arbitrary dimensions.

A.1.4. **The Hemispherical Transform.** The hemispherical transform $\mathcal{H}$, defined by $\mathcal{H}f(x) = \int_{\mathbb{S}^{d-1}} \mathbb{I}\{x'y \geq 0\}f(y)d\sigma(y)$, plays a central role in our analysis. It is a special case of the operator considered in the Funk-Hecke theorem above, with $F(t) = \mathbb{I}\{t \in [0,1]\}$, therefore the next proposition follows.

Notation. We define $\lambda(n,d) = \lambda_n\left(\mathbb{I}\left\{t \in [0,1]\right\}\right)$ for $d \geq 3$ and $\lambda(n,2) = \frac{2\sin(n\pi/2)}{n}$.

Proposition A.4. *When $d \geq 2$, the coefficients $\lambda(n,d)$ have the following expressions*

(*i*) $\lambda(0,d) = \frac{|\mathbb{S}^{d-1}|}{2}$

(*ii*) $\lambda(1,d) = \frac{|\mathbb{S}^{d-2}|}{d-1}$

(iii) $\forall p \in \mathbb{N} \setminus \{0\}, \ \lambda(2p, d) = 0$

(iv) $\forall p \in \mathbb{N}, \ \lambda(2p+1, d) = \frac{(-1)^p |\mathbb{S}^{d-2}| 1\cdot 3 \cdots (2p-1)}{(d-1)(d+1)\cdots(d+2p-1)}$.

Proof of Proposition A.4. Define $\alpha(n,d) := C_n^{\nu(d)}(1)|\mathbb{S}^{d-2}|^{-1}\lambda_n\left(\mathbb{I}\left\{t \in [0,1]\right\}\right)$. By the Funk-Hecke theorem

$$\alpha(n,d) = \int_0^1 C_n^{\nu(d)}(t)(1-t^2)^{(d-3)/2}dt,$$

thus using (A.7),

$$\alpha(n,d) = \frac{(-2)^{-n}(d-2)_n}{n!\left((d-1)/2\right)_n} \int_0^1 \frac{\mathrm{d}^n}{\mathrm{d}t^n}(1-t^2)^{n+(d-3)/2}dt.$$

Therefore for $n \geq 1$ and $d \geq 3$,

$$\alpha(n,d) = -\frac{(-2)^{-n}(d-2)_n}{n!\left((d-1)/2\right)_n} \left.\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}}(1-t^2)^{n-1+(d-3)/2}dt\right|_{t=0}$$

since the term on the right hand-side is equal to 0 for $t=1$. To prove that the coefficients $\alpha(2p,d)$ are equal to zero for p positive it is enough to prove

$$\left.\frac{\mathrm{d}^{2p+1}}{\mathrm{d}t^{2p+1}}(1-t^2)^{2p+1+m}\right|_{t=0} = 0, \quad \forall m \geq 1, \ p \geq 0.$$

The Faá di Bruno formula gives that this quantity is equal to

$$\left.\sum_{k_1+2k_2=2p+1} \frac{(-1)^{2p+1-k_2}(2p+1)!(m+1)\cdots(2p+1+m)}{k_1!k_2!}(1-t^2)^{m+k_2}(2t)^{k_1}\right|_{t=0}.$$

and the result follows since k_1 in the sum cannot be equal to 0.

When $n = 2p+1$ for $p \in \mathbb{N}$ we obtain, again using the Faá di Bruno formula, that the derivative at $t=0$ is equal to

$$(-1)^p \frac{(2p)!}{p!}\left[(2p+1+(d-3)/2)(2p+(d-3)/2)\cdots(p+2+(d-3)/2)\right].$$

Together with (A.9), the desired result follows. For the case $d=2$ we use Proposition A.1. □

Define $\mathrm{L}^2_{\text{odd}}(\mathbb{S}^{d-1})$ and $\mathrm{H}^s_{\text{odd}}(\mathbb{S}^{d-1})$ as the restrictions of $\mathrm{L}^2(\mathbb{S}^{d-1})$ and $\mathrm{H}^s(\mathbb{S}^{d-1})$ to odd functions and similarly $\mathrm{L}^2_{\text{even}}(\mathbb{S}^{d-1})$ and $\mathrm{H}^s_{\text{even}}(\mathbb{S}^{d-1})$ for even functions. The following corollary is a direct consequence of the Funk-Hecke Theorem and Proposition A.4, and corresponds to an observation made in Remark A.1 for the $d=2$ case.

Corollary A.1. *The null space of the hemispherical transform $\mathcal{H}$ is given by*

$$\ker\, \mathcal{H} = \bigoplus_{p=1}^{\infty} H^{2p,d} = \left\{ f \in \mathrm{L}^2_{\text{even}}(\mathbb{S}^{d-1}) : \int_{\mathbb{S}^{d-1}} f(x)d\sigma(x) = 0 \right\},$$

when $\mathcal{H}$ is viewed as an operator on $\mathrm{L}^2(\mathbb{S}^{d-1})$. *The spaces $H^{0,d}$ and $H^{2p+1,d}$ for $p \in \mathbb{N}$ are the eigenspaces associated with the non-zero eigenvalues of $\mathcal{H}$.*

As a consequence of Proposition A.4, $\mathcal{H}$ is not injective and restrictions have to be imposed in order to ensure identification of f_β. Section 3 presents sufficient conditions that allows us to reconstruct f_β from f_β^-.

The following proposition can be found in Rubin (1999).

Proposition A.5. *$\mathcal{H}$ is a bijection from* $\mathrm{L}^2_{\mathrm{odd}}(\mathbb{S}^{d-1})$ *to* $\mathrm{H}^{d/2}_{\mathrm{odd}}(\mathbb{S}^{d-1})$.

Lemma A.2.

$$h(n,d) \asymp n^{d-2}, \tag{A.26}$$

$$|\lambda(2p+1,d)| \asymp p^{-d/2}. \tag{A.27}$$

Proof. Estimate (A.26) is clearly satisfied when $d=2$ and 3 since $h(n,2)=2$ and $h(n,3)=2n+1$. When $d \geq 4$ we have

$$h(n,d) = \frac{2}{(d-2)!}(n+(d-2)/2)[(n+1)(n+2)\cdots(n+d-3)],$$

and the results follow.

Next we turn to (A.27). When d is even and $p \geq d/2$

$$|\lambda(2p+1,d)| = \frac{\kappa_d}{(2p+1)(2p+3)\cdots(2p+d-1)}$$

where

$$\kappa_d = \frac{|\mathbb{S}^{d-2}|1\cdot 3\cdots(d-1)}{d-1}$$

and (A.27) follows. Sterling's double inequality (see Feller (1968) p.50-53), that is,

$$\sqrt{2\pi}n^{n+1/2}\exp\left(-n+\frac{1}{12n+1}\right) < n! < \sqrt{2\pi}n^{n+1/2}\exp\left(-n+\frac{1}{12n}\right),$$

implies that

$$\frac{(2^p p!)^2}{(2p)!} \asymp \sqrt{p}$$

and therefore

$$1\cdot 3\cdots(2p-1) \asymp \sqrt{p}2\cdot 4\cdots(2p).$$

Thus for $p \geq d/2$ and d odd we have

$$|\lambda(2p+1,d)| \asymp \frac{\sqrt{p}}{(2p+2)(2p+4)\cdots(2p+d-1)}$$

and (A.27) holds for both even and odd d. □

We can now easily check that

Proposition A.6. *For all $s > 0$, there exists positive constants C_l and C_u such that for all f in* $\mathrm{H}^s(\mathbb{S}^{d-1})$

$$C_l \left\|f^-\right\|_{2,s} \leq \left\|\mathcal{H}(f^-)\right\|_{2,s+d/2} \leq C_u \left\|f^-\right\|_{2,s}.$$

Proof of Proposition A.6. By definition we have

$$\|\mathcal{H}\left(f^-\right)\|^2_{2,s+d/2} = \sum_{p=0}^{\infty}(1+\zeta_{2p+1,d})^{s+d/2}\|Q_{2p+1,d}\mathcal{H}(f^-)\|_2^2$$

where according to the Funk-Hecke Theorem

$$\begin{aligned} Q_{2p+1,d}\mathcal{H}(f^-) &= Q_{2p+1,d}\mathcal{H}\left(\sum_{q=0}^{\infty} Q_{2q+1,d}f\right) \\ &= Q_{2p+1,d}\left(\sum_{q=0}^{\infty}\lambda(2q+1,d)Q_{2q+1,d}f\right) \\ &= \lambda(2p+1,d)Q_{2p+1,d}f. \end{aligned}$$

The result follows since Lemma A.2 implies that $(1+\zeta_{2p+1,d})^{s+d/2}\lambda^2(2p+1,d) \asymp (1+\zeta_{2p+1,d})^s$. □

The factor $d/2$ in Proposition A.6 corresponds to the degree of "regularization" due to smoothing by $\mathcal{H}$. Now the inverse of an odd function f^- is given by

$$\mathcal{H}^{-1}(f^-)(y) = \sum_{p=0}^{\infty}\frac{1}{\lambda(2p+1,d)}\int_{\mathbb{S}^{d-1}} q_{2p+1,d}(x,y)f^-(x)d\sigma(x). \tag{A.28}$$

This is straightforward given our results at hand: for example, operate $\mathcal{H}$ on the RHS to see:

$$\begin{aligned} \mathcal{H}\left(\sum_{p=0}^{\infty}\frac{1}{\lambda(2p+1,d)}\int_{\mathbb{S}^{d-1}} q_{2p+1,d}(x,y)f^-(x)d\sigma(x)\right) &= \sum_{p=0}^{\infty}\frac{1}{\lambda(2p+1,d)}\mathcal{H}Q_{2p+1,d}f^- \\ &= \sum_{p=0}^{\infty}\frac{\lambda(2p+1,d)}{\lambda(2p+1,d)}Q_{2p+1,d}f^- \text{ (by the Funk-Hecke Theorem)} \\ &= f^-. \end{aligned}$$

If f^- belongs to $\mathrm{H}^{d/2}(\mathbb{S}^{d-1})$, then $\mathcal{H}^{-1}(f^-)(b)$ is a well-defined $\mathrm{L}^2(\mathbb{S}^{d-1})$ function. Otherwise it should be understood as a distribution and is only defined in a Sobolev space with negative exponent. Moreover, if d is a multiple of 4, it is possible to relate the inverse of the operator $\mathcal{H}$ with differentiation as in the case of $d = 2$:

Proposition A.7. *If d is a multiple of 4,*

$$\mathcal{H}^{-1} = |\mathbb{S}^{d-2}| \prod_{k=1}^{d/4} [-\Delta^S + 2(k-1)(d-2k)].$$

Proof of Proposition A.7. If we consider the case where d is even, we know from Proposition A.4, that

$$\frac{1}{\lambda(2p+1,d)} = (-1)^p |\mathbb{S}^{d-2}|(2p+1)(2p+3)\dots(d+2p-1).$$

Thus if d is a multiple of 4,

$$\frac{1}{\lambda(2p+1,d)} = |\mathbb{S}^{d-2}| \prod_{k=1}^{d/4} [-\zeta_{2p+1,d} + 2(k-1)(d-2k)].$$

Using this and (4.1),

$$\begin{aligned}\mathcal{H}^{-1} &= \sum_{p=0}^{\infty} \frac{1}{\lambda(2p+1,d)} Q_{2p+1,d} \\ &= \sum_{p=0}^{\infty} |\mathbb{S}^{d-2}| \left(\prod_{k=1}^{d/4} [-\zeta_{2p+1,d} + 2(k-1)(d-2k)] \right) Q_{2p+1,d}.\end{aligned}$$

Recall (A.18) and the proposition is proved. $\square$

This connection between the inverse of $\mathcal{H}$ and differentiation suggests that a Bernstein-type inequality might hold for $\mathcal{H}^{-1}$. Indeed, even though the above inversion formula is concerned with d's that are multiples of 4, the following Bernstein inequality holds for every dimension.

Theorem A.3 (Bernstein inequality)**.** *For every $d \geq 2$ and every $q \in [1, \infty]$, there exists a positive constant $B(d,q)$ such that for all P in $\bigoplus_{p=0}^{T} H^{2p+1,d}$,*

$$\|\mathcal{H}^{-1} P\|_q \leq B(d,q) T^{d/2} \|P\|_q. \tag{A.29}$$

Proof of Theorem A.3. We can write

$$\mathcal{H}^{-1} = P_1(D) - P_2(D)$$

where $P_1(D)$ and $P_2(D)$ are defined for all odd function f^- by

$$\begin{aligned}P_1(D) f^- &= \sum_{p=0}^{\infty} \frac{1}{\lambda(4p+3)} \int_{\mathbb{S}^{d-1}} q_{4p+3}(x,y) f^-(x) d\sigma(x) \\ P_2(D) f^- &= -\sum_{p=0}^{\infty} \frac{1}{\lambda(4p+1)} \int_{\mathbb{S}^{d-1}} q_{4p+1}(x,y) f^-(x) d\sigma(x).\end{aligned}$$

$P_1(D)$ and $P_2(D)$ are two unbounded operators on $B = \mathrm{L}^q_{\mathrm{odd}}(\mathbb{S}^{d-1})$ with non-positive eigenvalues. We apply Theorem 3.2. of Ditzian (1998) to $-P_1(D)$ and $-P_2(D)$ choosing $\alpha = 1$. Condition (1.6) of Ditzian (1998) can be verified using Proposition 2.2 with $r = 1$ and $p = q$ and the fact that for the Cesaro kernels C_h^l are uniformly bounded in $\mathrm{L}^1(\mathbb{S}^{d-1})$ for $l > \frac{d-2}{2}$ (see, e.g. Bonami and Clerc, 1973). We see, using the triangle inequality, that for all P in $\bigoplus_{p=0}^{T} H^{2p+1,d}$,

$$\begin{aligned}\|\mathcal{H}^{-1}P\|_q &\le C\frac{1}{\lambda^2(2T+1,d)}\|P\|_q\\ &\le CT^d\|P\|_q.\end{aligned}$$

The last inequality follows from (A.27). □

Rubin (1999) gives other inversion formulas for the Hemispherical transform in terms of differential operators. The fact that the inversion roughly corresponds to differentiation is another manifestation of the ill-posedness of our problem at hand. The inverse operator $\mathcal{H}^{-1}$ is indeed unbounded. We call the factor $d/2$ in (A.29) the degree of ill-posedness of the inverse problem. For the case $q = 2$, there exists a lower bound for $\|\mathcal{H}^{-1}P\|_q$ in (A.29) of order $T^{d/2}$ as well, implying that the upper bound $T^{d/2}$ in the order of T obtained in Theorem A.3 is tight.

A.1.5. **Estimators for the Choice Probability Function.** This section considers estimation of the choice probability function r and its extension R. We propose an estimator for r, which, in turn, yields a computationally simple estimator for f_β. Also the asymptotic results presented here are useful for the next section where we study the limiting properties of our estimator for the random coefficients density f_β.

Since R is square integrable on $\mathbb{S}^{d-1}$, it has a condensed harmonic expansion which enables us to obtain the expressions in the next theorem.

Theorem A.4. *For x in $\mathbb{S}^{d-1}$, we have*

$$R(x) = \frac{1}{2} + \sum_{p=0}^{\infty}\mathbb{E}\left[\frac{(2Y-1)}{f_X(X)}q_{2p+1,d}(X,x)\right]. \tag{A.30}$$

This suggests an estimator of the form $\hat{R}_1(x) = \frac{1}{2} + \hat{R}_1^-$ with

$$\hat{R}_1^-(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{(2y_i-1)}{\hat{f}_X(x_i)}\sum_{p=0}^{T_N} q_{2p+1,d}(x_i,x)$$

where $\hat{f}_X$ is an estimator of f_X and T_N is a suitably chosen sequence diverging to infinity with N. Note that the second summation corresponds to the Dirichlet kernel. We can generalize this, by introducing a class of estimators of the form

$$\hat{R}_2^-(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{(2y_i-1)}{\hat{f}_X(x_i)}K_{2T_N}^-(x_i,x) \tag{A.31}$$

where $K_{2T_N}^-$ is the odd part of a kernel of the form (A.23) satisfying Assumption A.1, such as the two kernels in Proposition A.3.

The estimator (A.31) is convenient, though the plug-in term $\hat{f}_X$ has to be treated with care. We avoid restrictive assumptions on the distributions of covariates and allow $f_X(x)$ to decay to zero as x approaches the boundary of its support H^+. To deal with the latter problem, we modify (A.31) by

$$\hat{R}^-(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{(2y_i-1)K_{2T_N}^-(x_i,x)}{\max\left(\hat{f}_X(x_i), m_N\right)} \tag{A.32}$$

where m_N is a trimming factor going to 0 with the sample size. Our estimator for R is then

$$\widehat{R} = \frac{1}{2} + \hat{R}^-. \tag{A.33}$$

Remark A.4. Alternative estimators of R^- are available. For example, one may use kernel regression on the sphere to estimate r in order to obtain an estimator for R^-. As noted before, however, we then need to use numerical integration to evaluate (4.5) to calculate $\hat{f}_\beta^-$.

Proof of Theorem A.4. R has the following condensed harmonic expansion

$$R(x) = \frac{1}{2} + \sum_{p=1}^{\infty}(Q_{2p+1,d}R)(x).$$

We then write using (3.2), changing variables and using (A.11),

$$\begin{aligned}
(Q_{2p+1,d}R)(x) &= \int_{\mathbb{S}^{d-1}} q_{2p+1,d}(x,z)R(z)d\sigma(z)\\
&= \int_{H^+} q_{2p+1,d}(x,z)r(z)d\sigma(z) + \int_{H^-} q_{2p+1,d}(x,z)(1-r(-z))d\sigma(z)\\
&= \int_{H^+} q_{2p+1,d}(x,z)r(z)d\sigma(z) - \int_{H^+} q_{2p+1,d}(x,z)(1-r(z))d\sigma(z)\\
&= \int_{H^+} q_{2p+1,d}(x,z)(2r(z)-1)d\sigma(z)\\
&= \int_{H^+} q_{2p+1,d}(x,z)\mathbb{E}\left[\left.\frac{2Y-1}{f_X(z)}\right| X=z\right] f_X(z)d\sigma(z)\\
&= \mathbb{E}\left[\frac{(2Y-1)q_{2p+1,d}(x,X)}{f_X(X)}\right].
\end{aligned}$$

$\square$

A.1.6. **Proofs of Main Results.**

Proof of Proposition 3.1. It is straightforward that the model (1.1) and Assumption 1.1 imply that the choice probability function r given by (1.2) is homogeneous of degree 0. Proposition A.5 along with the fact that $R = \frac{1}{2} + \mathcal{H}\left(f_\beta^-\right)$ with $f_\beta^- \in \mathrm{L}^2_{\mathrm{odd}}(\mathbb{S}^{d-1})$ implies that R belongs to $\mathrm{H}^{d/2}(\mathbb{S}^{d-1})$. We now turn to the proof of sufficiency. If the extension R given by (3.2) belongs to $\mathrm{H}^{d/2}(\mathbb{S}^{d-1})$ then so does R^- and Proposition A.5 shows that there exists a unique odd function f^- in $\mathrm{L}^2(\mathbb{S}^{d-1})$ such that

$$R = \frac{1}{2} + \mathcal{H}\left(f^-\right) = \mathcal{H}\left(\frac{1}{|\mathbb{S}^{d-1}|} + f^-\right).$$

Moreover, since $0 \le R(x) \le 1$ holds for every $x \in \mathbb{S}^{d-1}$, the above relationship implies that $\frac{1}{2} \ge \mathcal{H}f^-(x), \forall x \in \mathbb{S}^{d-1}$. But $\mathcal{H}f^-(x) \ge \int_{\{f^-(b)\ge 0\}} f^-(b)d\sigma(b)$ holds for some x. Therefore we conclude that $\frac{1}{2} \ge \int_{\{f^-(b)\ge 0\}} f^-(b)d\sigma(x) = -\int_{\{f^-(b)\le 0\}} f^-(b)d\sigma(b)$, thus $\int_{\mathbb{S}^{d-1}} |f^-(b)|d\sigma(b) \le 1$. Also, following the discussion in Section A.1.3, $\frac{1}{|\mathbb{S}^{d-1}|} + f^-$ integrates to 1. We have seen in Corollary A.1 that for even function g that has 0 as the coefficient of degree 0 in its expansion on the surface harmonics (i.e. an even function that integrates to zero over the sphere),

$$R = \mathcal{H}\left(g + \frac{1}{|\mathbb{S}^{d-1}|} + f^-\right)$$

holds. Now consider

$$g = |f^-| - \frac{1}{|\mathbb{S}^{d-1}|}\int_{\mathbb{S}^{d-1}} |f^-(b)|d\sigma(b),$$

then this certainly is even and integrates to zero. Using this, define

$$f_\beta^* := g + \frac{1}{|\mathbb{S}^{d-1}|} + f^- = 2f^-\mathbb{I}\{f^- > 0\} + \frac{1}{|\mathbb{S}^{d-1}|}\left(1 - \int_{\mathbb{S}^{d-1}} |f^-(b)|d\sigma(b)\right) \ge 0.$$

Obviously $f_\beta^{*-} = f^-$. This function f_β^* is non-negative and integrates to one, and thus it is a proper probability density function (pdf). It is indeed bounded from below by $\frac{1}{|\mathbb{S}^{d-1}|}\left(1 - \int_{\mathbb{S}^{d-1}} |f^-(b)|d\sigma(b)\right)$. As a consequence, there exists a pdf f_β^* such that

$$R = \mathcal{H}\left(f_\beta^*\right) = \frac{1}{2} + \mathcal{H}\left(f_\beta^{*-}\right)$$

and for all x in H^+, $r(x) = \mathcal{H}\left(f_\beta^*\right)(x)$. □

Proof of Theorem 4.1. We use the shorthand notation $\mathbb{I}(b) := \mathbb{I}\{f_\beta^-(b) > 0\}$ and $\hat{\mathbb{I}}(b) := \mathbb{I}\{\hat{f}_\beta^-(b) > 0\}$. Then $f_\beta = 2f_\beta^-\mathbb{I}$ and $\hat{f}_\beta = 2\hat{f}_\beta^-\hat{\mathbb{I}}$. We write

$$\overline{f}^-_{\beta,T}(b) = \frac{1}{N}\sum_{i=1}^N \frac{(2y_i - 1)\mathcal{H}^{-1}\left(K^-_{2T_N}(x_i, \cdot)\right)(b)}{\max\left(f_X(x_i), m_N\right)}$$

$$\overline{f}_\beta^{\,-}(b) = \frac{1}{N}\sum_{i=1}^{N}\frac{(2y_i-1)\mathcal{H}^{-1}\left(K^-_{2T_N}(x_i,\cdot)\right)(b)}{f_X(x_i)}$$

and use the decomposition

$$\hat{f}_\beta^- - f_\beta^- = \left(\hat{f}_\beta^- - \overline{f}_{\beta,T}^{\,-}\right) + \left(\overline{f}_{\beta,T}^{\,-} - \mathbb{E}\left[\overline{f}_{\beta,T}^{\,-}\right]\right) + \left(\mathbb{E}\left[\overline{f}_{\beta,T}^{\,-}\right] - \mathbb{E}\left[\overline{f}_{\beta}^{\,-}\right]\right) + \left(\mathbb{E}\left[\overline{f}_{\beta}^{\,-}\right] - f_\beta^-\right), \tag{A.34}$$

and denote the terms on the right hand side by S_{p} (stochastic component due to plug-in), S_{e} (stochastic component of the infeasible estimator $\overline{f}_{\beta,T}^{\,-}$), B_{t} (trimming bias) and B_{a} (approximation bias).

Take $q \in [1,\infty)$,

$$\begin{aligned}
\|\hat{f}_\beta - f_\beta\|_q^q &= \int (\hat{f}_\beta(b) - f_\beta(b))^q d\sigma(b)\\
&= \int_{\mathbb{I}(b)=1,\hat{\mathbb{I}}(b)=1} (\hat{f}_\beta(b) - f_\beta(b))^q d\sigma(b) + \int_{\mathbb{I}(b)=0,\hat{\mathbb{I}}(b)=1} (\hat{f}_\beta(b) - f_\beta(b))^q d\sigma(b)\\
&\quad + \int_{\mathbb{I}(b)=1,\hat{\mathbb{I}}(b)=0} (\hat{f}_\beta(b) - f_\beta(b))^q d\sigma(b) + \int_{\mathbb{I}(b)=0,\hat{\mathbb{I}}(b)=0} (\hat{f}_\beta(b) - f_\beta(b))^q d\sigma(b)\\
&:= A_1 + A_2 + A_3 + A_4.
\end{aligned}$$

Obviously

$$A_1 = \int_{\mathbb{I}(b)=1,\hat{\mathbb{I}}(b)=1} (2\hat{f}_\beta^-(b) - 2f_\beta^-(b))^q d\sigma(b)$$

and $A_4 = 0$. Also,

$$A_2 = \int_{\mathbb{I}(b)=0,\hat{\mathbb{I}}(b)=1} (2\hat{f}_\beta^-(b) - f_\beta(b))^q d\sigma(b).$$

But given $\mathbb{I}(b) = 0$ and $\hat{\mathbb{I}}(b) = 1$, $2\hat{f}_\beta^-(b) > 0$, $f_\beta(b) = 0$ and $2f_\beta^-(b) \le 0$, so replacing f_β with $2f_\beta^-$ in the bracket,

$$A_2 \le \int_{\mathbb{I}(b)=0,\hat{\mathbb{I}}(b)=1} (2\hat{f}_\beta^-(b) - 2f_\beta^-(b))^q d\sigma(b).$$

Similarly,

$$A_3 = \int_{\mathbb{I}(b)=1,\hat{\mathbb{I}}(b)=0} (\hat{f}_\beta(b) - 2f_\beta^-(b))^q d\sigma(b).$$

and given $\mathbb{I}(b) = 1$ and $\hat{\mathbb{I}}(b) = 0$, $2f_\beta^-(b) > 0$, $\hat{f}_\beta(b) = 0$ and $2\hat{f}_\beta^-(b) \le 0$, so replacing f_β with $2f_\beta^-$ in the bracket,

$$A_3 \le \int_{\mathbb{I}(b)=0,\hat{\mathbb{I}}(b)=1} (2\hat{f}_\beta^-(b) - 2f_\beta^-(b))^q d\sigma(b).$$

Overall,

$$\|\hat{f}_\beta - f_\beta\|_q^q \le 2^q \|\hat{f}_\beta^- - f_\beta^-\|_q^q.$$

A similar proof can be carried out replacing $\mathrm{L}^q(\mathbb{S}^{d-1})$ by $\mathrm{L}^\infty(\mathbb{S}^{d-1})$. Thus it is enough to consider the behavior of $\hat{f}_\beta^- - f_\beta^-$ instead of $\hat{f}_\beta - f_\beta$. As noted above, the former can be decomposed into four terms, S_{p}, S_{e}, B_{t} and B_{a}.

We start with the analysis of S_{p}. Note that for $q \in [1, \infty]$

$$\begin{aligned}
\|S_{\mathrm{p}}\|_q &= \left\| \mathcal{H}^{-1} \left(\frac{1}{N} \sum_{i=1}^N \frac{(2y_i - 1) K_{2T_N}^-(x_i, \cdot)}{\max(f_X(x_i), m_N)} \left(\frac{\max(f_X(x_i), m_N)}{\max\left(\hat{f}_X(x_i), m_N\right)} - 1 \right) \right) \right\|_q \\
&\le B(d,q) T_N^{d/2} \left\| \frac{1}{N} \sum_{i=1}^N \frac{(2y_i - 1) K_{2T_N}^-(x_i, \cdot)}{\max(f_X(x_i), m_N)} \left(\frac{\max(f_X(x_i), m_N)}{\max\left(\hat{f}_X(x_i), m_N\right)} - 1 \right) \right\|_q \quad \text{(by Theorem A.3)} \\
&\le B(d,q) T_N^{d/2} m_N^{-1} \left\| \frac{1}{N} \sum_{i=1}^N |K_{2T_N}(x_i, \cdot)| \right\|_q \max_{i=1,\dots,N} \left| \frac{\max(f_X(x_i), m_N)}{\max\left(\hat{f}_X(x_i), m_N\right)} - 1 \right| \\
&\le B(d,q) T_N^{d/2} m_N^{-2} \left\| \frac{1}{N} \sum_{i=1}^N |K_{2T_N}(x_i, \cdot)| \right\|_q \max_{i=1,\dots,N} \left| f_X(x_i) - \hat{f}_X(x_i) \right|
\end{aligned}$$

holds, where we have used the triangle inequality. The L^q-norm on the right hand side is bounded from above by

$$\left\| \frac{1}{N} \sum_{i=1}^N |K_{2T_N}(x_i, \cdot)| - \mathbb{E}\,|K_{2T_N}(X, \cdot)| \right\|_q + \left\| \mathbb{E}\,|K_{2T_N}(X, \cdot)| \right\|_q := \|T_1\|_q + \|T_2\|_q. \tag{A.35}$$

First consider the term $\|T_1\|_q$. We begin with the case of $q \in [1, 2]$. By the Hölder inequality,

$$\begin{aligned}
\mathbb{E}\left[\|T_1\|_q^q\right] &= \int_{\mathbb{S}^{d-1}} \mathbb{E}\left[T_1(x)^q\right] d\sigma(x) \\
&\le \int_{\mathbb{S}^{d-1}} \mathbb{E}\left[T_1(x)^2\right]^{q/2} d\sigma(x)
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}\left[T_1(x)^2\right] &\le \frac{1}{N} \mathbb{E}\left[\left(K_{2T_N}(X, x)\right)^2\right] \\
&\le \frac{C}{N} \|K_{2T_N}(\star_2, x)\|_2^2 \quad \text{(boundedness assumption on } f_X\text{)} \\
&= \frac{C}{N} \left\| \sum_{n=0}^{2T_N} \chi(n, 2T_N) q_{n,d}(\star_2, x) \right\|_2^2 \\
&\le \frac{C}{N} \sum_{n=0}^{2T_N} \|q_{n,d}(\star_2, x)\|_2^2 \quad \text{(by Assumption A.1(iv))}
\end{aligned} \tag{A.36}$$

$$\leq \frac{C}{N} \sum_{n=0}^{2T_N} \frac{h^2(n,d) \left\| C_n^{\nu(d)}(\star_2' x) \right\|_2^2}{|\mathbb{S}^{d-1}|^2 (C_n^{\nu(d)}(1))^2}$$

$$\leq \frac{C}{N} \sum_{n=0}^{2T_N} h(n,d) \quad \text{(by (A.12))}$$

$$\leq \frac{C T_N^{d-1}}{N} \quad \text{(by Lemma A.2)}.$$

By the Markov inequality,

$$T_N^{d/2} m_N^{-2} \|T_1\|_q = O_p\left(m_N^{-2} N^{-1/2} T_N^{(2d-1)/2} \right), \tag{A.37}$$

providing a convergence rate for $\|T_1\|_q, q \in [1,2]$. So if we can establish a similar rate for $\|T_1\|_\infty$, all $\mathrm{L}^q(\mathbb{S}^{d-1})$ convergence rates of T_1 for $q \in (2,\infty]$ can be interpolated between the $\mathrm{L}^2(\mathbb{S}^{d-1})$ and $\mathrm{L}^\infty(\mathbb{S}^{d-1})$ convergence rates using the following inequality:

$$\forall f \in \mathrm{L}^\infty(\mathbb{S}^{d-1}),\ \|f\|_q \leq \|f\|_2^{2/q} \|f\|_\infty^{1-2/q}. \tag{A.38}$$

To see this, note

$$\begin{aligned} \|f\|_q &= \|f^2 |f|^{q-2}\|_1^{1/q} \\ &\leq \left[\|f^2\|_1 \||f|^{q-2}\|_\infty \right]^{1/q} \quad \text{(by Hölder)} \\ &= \|f\|_2^{2/q} \|f\|_\infty^{1-2/q}. \end{aligned}$$

We can thus focus on $\|T_1\|_\infty$. We cover the sphere $\mathbb{S}^{d-1}$ by $\mathfrak{N}(N,d)$ geodesic balls (caps) $(B_i)_{i=1}^{\mathfrak{N}(N,d)}$ of centers $(\tilde{x}_i)_{i=1}^{\mathfrak{N}(N,d)}$ and radius $R(N,d)$, that is, $B_i = \{x \in \mathbb{S}^{d-1} :\ \|x - \tilde{x}_i\| \leq R(N,d)\}$. As the notation suggests, we let the radius of the balls depend on N and d, as specified more precisely below. Note that $\mathfrak{N}(N,d) \asymp R(N,d)^{-(d-1)}$.

We now prove that for every $\epsilon > 0$ positive, there exists a positive M such that

$$\mathbb{P}\left(v_N T_N^{d/2} m_N^{-2} \sup_{x \in \mathbb{S}^{d-1}} |T_1(x)| \geq M \right) \leq \epsilon \tag{A.39}$$

holds for an appropriately chosen sequence $v_N \uparrow \infty$. Write

$$\begin{aligned} &\mathbb{P}\left(v_N T_N^{d/2} m_N^{-2} \sup_{x \in \mathbb{S}^{d-1}} |T_1(x)| \geq M \right) \\ &\leq \mathbb{P}\left(\bigcup_{i=1,\dots,\mathfrak{N}(N,d)} \left\{ v_N T_N^{d/2} m_N^{-2} |T_1(\tilde{x}_i)| \geq M/2 \right\} \right) \end{aligned} \tag{A.40}$$

$$+\,\mathbb{P}\left(\exists i\in\{1,\dots,\mathfrak{N}(N,d)\}:\ v_N T_N^{d/2} m_N^{-2}\sup_{x\in B_i}|T_1(x)-T_1(\tilde{x}_i)|\geq M/2\right)$$
$$\leq \mathfrak{N}(N,d)\sup_{i=1,\dots,\mathfrak{N}_N}\mathbb{P}\left(v_N T_N^{d/2} m_N^{-2}|T_1(\tilde{x}_i)|\geq M/2\right)$$

where the last inequality is obtained using Assumption A.1 (ii) on the kernel and letting $R(N,d)\asymp m_N^2 v_N^{-1} T_N^{-(d/2+\alpha)}M$ (where α is given in Assumption A.1 (ii)). Notice

$$\mathbb{P}\left(v_N T_N^{d/2} m_N^{-2}|T_1(\tilde{x}_i)|\geq M/2\right) \tag{A.41}$$
$$=\mathbb{P}\left(\left|\sum_{j=1}^{N}\frac{|K_{2T_N}(x_j,\tilde{x}_i)|}{T_N^{d-1}}-\mathbb{E}\left[\frac{|K_{2T_N}(X,\tilde{x}_i)|}{T_N^{d-1}}\right]\right|\geq T_N^{-(d-1)}v_N^{-1}T_N^{-d/2}m_N^2 NM/2\right)$$
$$\leq 2\exp\left\{-\frac{1}{2}\left(\frac{t^2}{\omega+Lt/3}\right)\right\}\quad\text{(Bernstein inequality)}$$

where

$$t=T_N^{-(d-1)}v_N^{-1}T_N^{-d/2}m_N^2 NM/2$$
$$\omega\geq\sum_{j=1}^{N}\operatorname{var}\left(\frac{|K_{2T_N}(X_j,\tilde{x}_i)|}{T_N^{d-1}}\right)$$
$$\forall j=1,\dots,N,\ \left|\frac{K_{2T_N}(X_j,\tilde{x}_i)}{T_N^{d-1}}\right|\leq L\quad\text{(using (A.17) and (A.8))}.$$

The bound L in the last line is obtained by noting that $|K_{2T_N}(X_j,\tilde{x}_i)|=\left|\sum_{n=0}^{2T_N}\chi(n,2T_N)q_{n,d}(X_j,\tilde{x}_i)\right|\leq C\sum_{n=0}^{2T_N}|h(n,d)|\asymp T_N^{d-1}$, which follows from (A.17), (A.8) and (A.26). Here we can take $\omega=CN\mathbb{E}[K_{2T_N}(X,\tilde{x}_i)^2]$, then by the calculations in (A.36), we can write $\omega=CNT_N^{-(d-1)}$. ω is the leading term in the denominator of the exponent in the last inequality.

If we take $v_N=(\log N)^{-1/2}m_N^2 N^{1/2}T_N^{-(2d-1)/2}$, then

$$\frac{t^2}{\omega+Lt/3}\asymp(\log N)M^2. \tag{A.42}$$

Also, use this v_N in our choice of $R(N,d)$ made above to get:

$$R(N,d)\asymp m_N^2 v_N^{-1}T_N^{-(d/2+\alpha)}M=(\log(N))^{1/2}N^{-1/2}T_N^{\frac{d-1}{2}-\alpha}M$$

Thus

$$\mathfrak{N}(N,d)\asymp R(N,d)^{-(d-1)}=\exp\left(C_1\log N+o(\log N)\right) \tag{A.43}$$

for some constant C_1 that might be greater than $\frac{1}{2}(d-1)$, depending on the value of α. Indeed, T_N does not grow more than polynomially fast in N. (A.40), (A.41), (A.42) and (A.43) imply that, for a positive constants C and C_2,

$$\mathbb{P}\left(v_N T_N^{d/2} m_N^{-2} \sup_{x\in\mathbb{S}^{d-1}} |T_1(x)| \geq M\right) \leq C\exp\left\{(\log N)(C_1 - C_2M^2)\right\} \tag{A.44}$$

holds. For a large enough M, $C_1 - C_2M^2 < 0$ and the right hand side of (A.44) converges to zero, so (A.39) follows. In summary, we have just shown that

$$T_N^{d/2} m_N^{-2}\|T_1\|_\infty = O_p\left((\log N)^{1/2} m_N^{-2} N^{-1/2} T_N^{(2d-1)/2}\right)$$

and with (A.37) and (A.38) we also conclude that

$$T_N^{d/2} m_N^{-2}\|T_1\|_q = O_p\left((\log N)^{1/2-1/q} m_N^{-2} N^{-1/2} T_N^{(2d-1)/2}\right).$$

Concerning $\|T_2\|_q$, $q\in[1,\infty]$, since f_X is bounded by assumption, there exists a positive C such that

$$\|T_2\|_q \leq C\left\|\|K_{2T_N}(\star_1,\star_q)\|_1\right\|_q$$

where integration in $\|\cdot\|_1$ is with respect to argument $\star_1$ and integration in $\|\cdot\|_q$ is with respect to $\star_q$. But $\|K_{2T_N}(\star_1,\star_q)\|_1$ is a constant and does not depend on $\star_q$, as previously noted. Thus

$$\left\|\|K_{2T_N}(\star_1,\star_q)\|_1\right\|_q = |\mathbb{S}^{d-1}|^{1/q}\|K_{2T_N}(\star_1,\star_q)\|_1$$

and we conclude that this term is $O(1)$ using Assumption A.1 (i) on the kernel, thus

$$T_N^{d/2} m_N^{-2}\|T_2\|_q = O\left(m_N^{-2} T_N^{d/2}\right).$$

Analogously to our treatment of $\|T_1\|_q$, we can prove that when $q\in[1,2]$,

$$\|S_{\mathrm{e}}\|_q = O_p\left(m_N^{-1} N^{-1/2} T_N^{(2d-1)/2}\right),$$

while for $q\in(2,\infty]$

$$\|S_{\mathrm{e}}\|_q = O_p\left(m_N^{-1}(\log N)^{1/2-1/q} N^{-1/2} T_N^{(2d-1)/2}\right).$$

Let us now turn to the bias term induced by trimming

$$\begin{aligned} B_{\mathrm{t}}(b) &= \mathbb{E}\left[\frac{(2Y-1)\mathcal{H}^{-1}\left(K_{2T_N}^{-}(X,\cdot)\right)(b)}{f_X(X)}\left(\frac{f_X(X)}{\max(f_X(X),m_N)} - 1\right)\right] \\ &= \int_{\{z\in\mathbb{S}^{d-1}:\ f_X(z)<m_N\}} \mathbb{E}[2Y-1|X=z]\mathcal{H}^{-1}\left(K_{2T_N}^{-}(z,\cdot)\right)(b)\left(f_X(z)m_N^{-1} - 1\right)d\sigma(z). \end{aligned}$$

This yields

$$\begin{aligned} |B_{\mathrm{t}}(b)| &\leq \int_{\mathbb{S}^{d-1}} \left|\mathcal{H}^{-1}\left(K_{2T_N}^{-}(z,\cdot)\right)(b)\right| \mathbb{I}\left\{z\in\mathbb{S}^{d-1}:\ f_X(z)<m_N\right\}d\sigma(z) \\ &= \int_{\mathbb{S}^{d-1}} \left|\mathcal{H}^{-1}\left(K_{2T_N}^{-}(b,\cdot)\right)(z)\right| \mathbb{I}\left\{z\in\mathbb{S}^{d-1}:\ f_X(z)<m_N\right\}d\sigma(z) \quad \text{(using the condensed Harmonic expansion)}, \end{aligned}$$

thus, for every $1 \leq r \leq q$,

$$\begin{aligned}\|B_{\rm t}\|_q &\leq \left\|\mathcal{H}^{-1}\left(K_{2T_N}^{-}(b,\cdot)\right)\right\|_r \sigma\left(f_X < m_N\right)^{1/q-1/r+1} \quad \text{(from Proposition A.2)}\\ &\leq CB(d,r)T_N^{d/2+(d-1)(1-1/r)}\sigma\left(f_X < m_N\right)^{1/q-1/r+1}\end{aligned}$$

where in the last inequality we use Theorem A.3 and calculate an upper bound on the L^r-norm of the kernel by interpolation, using Hölder's inequality, between the uniformly bounded L^1-norm and the upper bound on the sup norm of the order of T_N^{d-1} seen previously, C is a constant. We finally treat $B_{\rm a}$ using Assumption A.1 (iii) with the condition that $f_\beta^- \in \mathrm{W}_q^s(\mathbb{S}^{d-1})$:

$$\|B_{\rm a}\|_q \leq CT_N^{-s}.$$

In the case where $f_X \geq m\ \sigma\ a.e.$, we use the decomposition

$$\begin{aligned}\hat{f}_\beta^- - f_\beta^- &= \left(\hat{f}_\beta^- - \overline{f}_\beta^-\right) + \left(\overline{f}_\beta^- - \mathbb{E}\left[\overline{f}_\beta^-\right]\right) + \left(\mathbb{E}\left[\overline{f}_\beta^-\right] - f_\beta^-\right)\\ &= \tilde{S}_{\rm p} + \tilde{S}_{\rm e} + B_{\rm a}.\end{aligned}$$

Now for example,

$$\|\tilde{S}_{\rm p}\|_q \leq B(d,q)T_N^{d/2}\left\|\frac{1}{N}\sum_{i=1}^N |K_{2T_N}(x_i,\cdot)|\right\|_q \frac{\max_{i=1,\dots,N}\left|f_X(x_i) - \hat{f}_X(x_i)\right|}{\min_{i=1,\dots,N}|\hat{f}_X(x_i)|},$$

because $\hat{f}_X$ is a consistent estimator in sup norm,

$$\forall \epsilon > 0,\ \exists N_0 > 0 :\ \forall n \geq N_0,\ \mathbb{P}\left(\min_{i=1,\dots,N}|\hat{f}_X(x_i)| > \frac{m}{2}\right) \leq \frac{\epsilon}{2},$$

and we can treat the terms $\tilde{S}_{\rm p}$ and $\tilde{S}_{\rm e}$ on this event. □

Proof of the corollaries 4.1, 4.2 and 4.3. The rate γs in Corollary 4.1 comes from the fact that it coincides with the maximum of

$$\min\left(\gamma s, -\gamma\frac{d}{2} - \rho + \frac{1}{2} - \gamma\frac{d-1}{2}, -\gamma\frac{d}{2} + r_X - 2\rho, -\gamma\frac{d}{2} + \rho\tau - \gamma(d-1)(1-1/q)\right). \tag{A.45}$$

for $r_X/2 \leq \rho < 1/2$ and $0 < \gamma < 1/(d-1)$ which is what we get from (4.9) and (4.10). Indeed, it is enough to find $\gamma(\rho)$ as the minimum of

$$\min\left(\gamma\left(s + \frac{d}{2}\right), -\rho + \frac{1}{2} - \gamma\frac{d-1}{2}, r_X - 2\rho, \rho\tau - \gamma(d-1)(1-1/q)\right). \tag{A.46}$$

The first is an increasing function of γ while the second and fourth are decreasing. The rest follows by simple computations. The proofs of the convergence in probability on Corollaries 4.2 and 4.3 is similar and simpler because there is only one parameter γ. In order to prove the strong uniform consistency in Corollary 4.1, noticing that the bias terms $B_{\rm t}$ and $B_{\rm a}$ are not stochastic and bounded after proper scaling, we just have to focus on $S_{\rm p}$ and $S_{\rm e}$ appearing in the proof of Theorem 4.1. Concerning $S_{\rm p}$, proceed as before and note that taking M large enough so that $C_1 - C_2M^2 < -1$ implies summability of the left hand side in (A.44). We conclude from the first Borel-Cantelli lemma that the probability that the events occur infinitely often is zero thus with probability one

$$\overline{\lim}_{N\to\infty} v_N B(d,\infty) T_N^{d/2} m_N^{-2} \sup_{x\in\mathbb{S}^{d-1}} |T_1(x)| < M.$$

The term T_2 is non-stochastic and its treatment in our previous analysis remains valid, therefore we can use the same non-stochastic upper bound. We then use Assumption 4.2 (iii) instead of Assumption 4.2 (ii) to show almost sure uniform boundedness of $S_{\rm p}$ after proper rescaling. The treatment of $S_{\rm e}$ is analogous to that of T_1. The proof is the same in Corollaries 4.2 and 4.3. □

Proof of Theorem 4.2. We first prove that the Lyapounov condition holds: there exists $\delta > 0$ such that for N going to infinity,

$$\frac{\mathbb{E}\left[\left|Z_N(b) - \mathbb{E}\left[Z_N(b)\right]\right|^{2+\delta}\right]}{N^{\delta/2}\left(\mathrm{var}\left(Z_N(b)\right)\right)^{1+\delta/2}} \to 0 \tag{A.47}$$

(see, e.g. Billingsley, 1995). We start from deriving a lower bound on $\mathrm{var}\left(Z_N(b)\right)$. Since $\mathbb{E}[Z_N(b)]$ converges to $f_\beta^-(b)$, it is enough to obtain a lower bound on

$$\begin{aligned}
\mathbb{E}[Z_N^2](b) &= 4\int_{H^+}\left(\sum_{p=0}^{T_N-1}\chi(2p+1,2T_N)\frac{q_{2p+1,d}(z,b)}{\max\left(f_X(z),m_N\right)\lambda(2p+1,d)}\right)^2 f_X(z)d\sigma(z)\\
&= 4\int_{H^+}\left(\sum_{p=0}^{T_N-1}\chi(2p+1,2T_N)\frac{q_{2p+1,d}(z,b)}{\lambda(2p+1,d)}\right)^2\left(\frac{1}{f_X(z)}\mathbb{I}\{f_X\ge m_N\}+f_X(z)m_N^{-2}\mathbb{I}\{f_X<m_N\}\right)d\sigma(z)\\
&\ge 4\frac{1}{\|f_X\|_\infty}\int_{H^+}\left(\sum_{p=0}^{T_N-1}\chi(2p+1,2T_N)\frac{q_{2p+1,d}(z,b)}{\lambda(2p+1,d)}\right)^2 d\sigma(z)\\
&\qquad -4\frac{1}{\|f_X\|_\infty}\int_{\{f_X<m_N\}}\left(\sum_{p=0}^{T_N-1}\chi(2p+1,2T_N)\frac{q_{2p+1,d}(z,b)}{\lambda(2p+1,d)}\right)^2 d\sigma(z)
\end{aligned}$$

With similar computations as (A.36), using as well (A.27), we know that there exists a constant C such that

$$\left\|\sum_{p=0}^{T_N-1}\chi(2p+1,2T_N)\frac{q_{2p+1,d}(z,\star)}{\lambda(2p+1,d)}\right\|_2 \le CT_N^{2d-1},$$

therefore using Proposition A.2 with $p=q=r=1$ we obtain

$$\mathbb{E}[Z_N^2](b) \ge \frac{4}{\|f_X\|_\infty}\sum_{p=0}^{T_N-1}\chi(2p+1,2T_N)^2\int_{H^+}\frac{q_{2p+1,d}(z,b)^2}{\lambda(2p+1,d)^2}d\sigma(z) - CT_N^{2d-1}\sigma\left(f_X<m_N\right).$$

Using Assumption A.1 (iv), the first term on the right hand side can be bounded from below by

$$C \sum_{p=0}^{\lfloor (T_N-1)/2 \rfloor} \left\| \frac{q_{2p+1,d}(z,b)}{\lambda(2p+1,d)} \right\|_2^2$$

i.e. by CT_N^{2d-1}. Thus as m_N decays to zero, $\sigma\left(f_X < m_N\right)$ decays to zero and

$$\mathbb{E}[Z_N^2](b) \geq CT_N^{2d-1}. \tag{A.48}$$

We now derive an upper bound of $\mathbb{E}\left[\left|Z_N(b)\right|^{2+\delta}\right]$ using Theorem A.3 and interpolation between $\mathrm{L}^\infty(\mathbb{S}^{d-1})$ and $\mathrm{L}^1(\mathbb{S}^{d-1})$ norms of the kernels using the Hölder inequality:

$$\begin{aligned}
\mathbb{E}\left[\left|Z_N\right|^{2+\delta}\right] &\leq \|f_X\|_\infty m_N^{-(2+\delta)} \left\|\mathcal{H}^{-1}\left(K_{2T_N}^-(z,\cdot)\right)\right\|_{2+\delta}^{2+\delta} \\
&\leq \|f_X\|_\infty m_N^{-(2+\delta)} B(d,2+\delta)^{2+\delta} T_N^{d(2+\delta)/2} \left\|K_{2T_N}^-(z,\cdot)\right\|_{2+\delta}^{2+\delta} \\
&\leq C m_N^{-(2+\delta)} T_N^{d(2+\delta)/2} T_N^{(d-1)(1+\delta)}.
\end{aligned}$$

By this and (A.48) an upper bound for the ratio appearing in (A.47) is given by

$$m_N^{-(2+\delta)} \left(\frac{T_N^{d-1}}{N}\right)^{\delta/2}.$$

Therefore the Lyapounov condition is satisfied if (4.20) holds, and it follows that $N^{1/2} s_N^{-1}(b) S_{\mathrm{e}} \xrightarrow{d} N(0,1)$.

We now need to prove that the remaining terms S_{p}, B_{t} and B_{a}, multiplied by $N^{1/2}s_N^{-1}(b)$, are $o_p(1)$. The term S_{p} is treated in a similar manner as in the proof of Theorem 4.1.

$$|S_{\mathrm{p}}(b)| \leq 2\left(\frac{1}{N}\sum_{i=1}^N \frac{\left|\mathcal{H}^{-1}\left(K_{2T_N}^-(x_i,\cdot)\right)(b)\right|}{\max(f_X(x_i), m_N)}\right) \max_{i=1,\dots,N}\left|\frac{\max\left(f_X(x_i), m_N\right)}{\max\left(\hat{f}_X^N(x_i), m_N\right)} - 1\right|.$$

Using the Markov inequality, the empirical average in the parenthesis is of the stochastic order of

$$m_N^{-1}\left\|\mathcal{H}^{-1}\left(K_{2T_N}^-(\star,\cdot)\right)\right\|_1.$$

But

$$\begin{aligned}
m_N^{-1}\left\|\mathcal{H}^{-1}\left(K_{2T_N}^-(\star,\cdot)\right)\right\|_1 &\leq B(d,1) T_N^{d/2} m_N^{-1}\left\|K_{2T_N}^-(\star,\cdot)\right\|_1 \\
&\leq B(d,1) T_N^{d/2} m_N^{-1}\left\|K_{2T_N}(\star,\cdot)\right\|_1
\end{aligned}$$

where the first inequality follows from Theorem A.3 and the second is obtained using the definition of the odd part and the triangle inequality. Note that the term $\left\|K_{2T_N}(\star,\cdot)\right\|_1$ in the last line does not depend on $\cdot$ and is uniformly bounded. By the lower bound (A.48) it is enough to show $N^{1/2}B(d,1)T_N^{-(d-1/2)}|S_{\mathrm{p}}(b)| = o_p(1)$. From the inequality above,

$$N^{1/2}B(d,1)T_N^{-(d-1/2)}|S_{\mathrm{p}}(b)| \leq \left(N^{1/2}T_N^{-(d-1)/2}m_N^{-1}\right)\max_{i=1,\dots,N}\left|\frac{\max\left(f_X(x_i),m_N\right)}{\max\left(\hat{f}_X(x_i),m_N\right)}-1\right|.$$

Its right hand side is of $o_p(1)$ if

$$\max_{i=1,\dots,N}\left|f_X(x_i)-\hat{f}_X(x_i)\right| = o_p\left(N^{-1/2}T_N^{(d-1)/2}m_N^2\right),$$

which is met under (4.19).

Let us now consider the bias term induced by the trimming procedure. In the proof of Theorem 4.1 we have obtained an upper bound for $\|B_{\mathrm{t}}\|_\infty$ and we deduce that

$$N^{1/2}T_N^{-(d-1/2)}\|B_{\mathrm{t}}\|_\infty = o(1)$$

when condition (4.22) is satisfied. Finally, $N^{1/2}T_N^{-(d-1/2)}\|B_{\mathrm{a}}\|_\infty = o(1)$ if condition (4.21) is satisfied. We conclude that the asymptotic normality holds for b such that $f_\beta(b) > 0$. The factor 4 in the variance comes from the fact that $\hat{f}_\beta = 2\hat{f}_\beta^{-}\hat{\mathbb{I}}$. $\square$

The proof of Theorem 4.3 is almost the same.

CREST (ENSAE), 3 avenue Pierre Larousse, 92245 Malakoff Cedex, France.

E-mail address: eric.gautier@ensae-paristech.fr

Cowles Foundation for Research in Economics, Yale University, New Haven, CT-06520.

E-mail address: yuichi.kitamura@yale.edu