Technical Supplement: Robust functional estimation using the median and spherical principal components

BY DANIEL GERVINI

Department of Mathematical Sciences, University of Wisconsin – Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, U.S.A.

gervini@uwm.edu

PROOF OF THEOREM 1

The proof is conditional on the X_is, so let us assume that $X_1, \ldots, X_n$ are non-random continuous functions. Since $x_{ij}x_{kj} = X_i(t_j)X_k(t_j) + X_i(t_j)\varepsilon_{kj} + X_k(t_j)\varepsilon_{ij} + \varepsilon_{ij}\varepsilon_{kj}$, we can write $\hat{G}_{ik} = A_m + B_m + C_m + D_m - \hat{\sigma}_i^2\,\delta_{ik}$, where

$$
\begin{aligned}
A_m &= \frac{1}{2}\sum_{j=1}^{m-1}\{X_i(t_j)X_k(t_j) + X_i(t_{j+1})X_k(t_{j+1})\}(t_{j+1}-t_j),\\
B_m &= \frac{1}{2}\sum_{j=1}^{m-1} X_i(t_j)(t_{j+1}-t_j)\varepsilon_{kj} + \frac{1}{2}\sum_{j=1}^{m-1} X_i(t_{j+1})(t_{j+1}-t_j)\varepsilon_{k,j+1},\\
C_m &= \frac{1}{2}\sum_{j=1}^{m-1} X_k(t_j)(t_{j+1}-t_j)\varepsilon_{ij} + \frac{1}{2}\sum_{j=1}^{m-1} X_k(t_{j+1})(t_{j+1}-t_j)\varepsilon_{i,j+1},\\
D_m &= \frac{1}{2}\sum_{j=1}^{m-1}(t_{j+1}-t_j)\varepsilon_{ij}\varepsilon_{kj} + \frac{1}{2}\sum_{j=1}^{m-1}(t_{j+1}-t_j)\varepsilon_{i,j+1}\varepsilon_{k,j+1}.
\end{aligned}
$$

By continuity of the X_is, for each j there exists $s_j \in [t_j, t_{j+1}]$ such that $\int_{t_j}^{t_{j+1}} X_i(t)\, X_k(t)dt = X_i(s_j)X_k(s_j)(t_{j+1}-t_j)$. Moreover, the X_is are uniformly continuous in T, so given $\varepsilon > 0$ there is a $\delta > 0$ such that $|s-t| < \delta$ implies $|X_i(s)X_i(s) - X_i(t)X_i(t)| < \varepsilon$. Since $\max_{1\leq j\leq m-1}(t_{j+1}-t_j) = o(1)$ as $m \to \infty$, there is an m_0 such that $(t_{j+1}-t_j) < \delta$ for any $m \geq m_0$ and therefore $|A_m - G_{ik}| < \varepsilon$ for any $m \geq m_0$. Hence $A_m \to G_{ik}$ as $m \to \infty$.

To see that $B_m \longrightarrow 0$ in probability, note that $E(B_m) = 0$ and

$$\mathrm{var}(B_m) \leq \max_{1 \leq j \leq m-1} (t_{j+1} - t_j) \frac{1}{2} \sup_{t \in T} |X_i(t)| \sigma_i^2 (t_m - t_1),$$

which goes to zero as $m \to \infty$. The proof that $C_m \longrightarrow 0$ in probability is analogous. Regarding D_m, we have $E(D_m) = \sigma_i^2 \delta_{ik}$ and

$$\mathrm{var}(D_m) \leq \frac{1}{2} \max_{1 \leq j \leq m-1} (t_{j+1} - t_j) \{E(\varepsilon_{ij}^4) E(\varepsilon_{kj}^4)\}^{1/2} (t_m - t_1),$$

which goes to zero as $m \to \infty$, so $D_m \longrightarrow \sigma_i^2 \delta_{ik}$ in probability and this completes the proof.

PROOF OF THEOREM 2

The proof is a straightforward application of Theorem 1 of Huber (1967), taking as parameter space Θ the $L^2(T)$ space endowed with the weak topology, which is then locally compact (see e.g. Massé, 1997, p. 142). Almost sure convergence in the weak topology of $L^2(T)$ means that $\langle \tilde{\mu}_n, f \rangle \to \langle M(P_0), f \rangle$ for all $f \in L^2(T)$, as stated. Note that Assumption (A-4) in Huber (1967) is satisfied if P_0 is not concentrated on a straight line, because $M(P_0)$ is then unique.

PROOF OF THEOREM 3

By symmetry of Z_i, for $i \neq j$ we have

$$\begin{aligned} \Omega_{ij} &= E\{Z_i Z_j / \sum_{k=1}^{p} \lambda_k Z_k^2\} \\ &= E[(-Z_i) Z_j / \{\sum_{k \neq i} \lambda_k Z_k^2 + \lambda_i (-Z_i)^2\}] \\ &= -\Omega_{ij}, \end{aligned}$$

so $\Omega_{ij} = 0$ and then Ω is diagonal. To see that the $\tilde{\lambda}_k$s are decreasing, note that $g_k(\lambda) = \lambda Z_k^2/(\sum_{j \neq k} \lambda_j Z_j^2 + \lambda Z_k^2)$ is increasing as a function of λ, strictly so if $Z_k \neq 0$. Therefore $\tilde{\lambda}_k = E\{g_k(\lambda_k)\} \geq E\{g_k(\lambda_{k+1})\}$, with strict inequality if $\lambda_k > \lambda_{k+1}$ and $\mathrm{pr}\{Z_k = 0\} < 1$. But

$$\begin{aligned} E\{g_k(\lambda_{k+1})\} &= E\left\{\frac{\lambda_{k+1} Z_k^2}{\sum_{j \neq k,k+1} \lambda_j Z_j^2 + \lambda_{k+1} Z_{k+1}^2 + \lambda_{k+1} Z_k^2}\right\} \\ &= E\left\{\frac{\lambda_{k+1} Z_{k+1}^2}{\sum_{j \neq k,k+1} \lambda_j Z_j^2 + \lambda_{k+1} Z_k^2 + \lambda_{k+1} Z_{k+1}^2}\right\} \end{aligned}$$

by exchangeability of Z_k with Z_{k+1}, and since $\lambda_{k+1} \leq \lambda_k$,

$$\begin{aligned} E\left\{\frac{\lambda_{k+1} Z_{k+1}^2}{\sum_{j \neq k,k+1} \lambda_j Z_j^2 + \lambda_{k+1} Z_k^2 + \lambda_{k+1} Z_{k+1}^2}\right\} &\geq E\left\{\frac{\lambda_{k+1} Z_{k+1}^2}{\sum_{j=1}^{p} \lambda_j Z_j^2}\right\} \\ &= \tilde{\lambda}_{k+1}, \end{aligned}$$

as claimed.

PROOF OF THEOREM 4

(i) Let $\varepsilon = P(\{z\})$ and write $P = (1-\varepsilon)Q + \varepsilon\delta_z$ for an appropriate Q. Then $F_P(y) = (1-\varepsilon)F_Q(y) + \varepsilon(\|z-y\| - \|z\|)$ and $F_P(z) = (1-\varepsilon)F_Q(z) - \varepsilon\|z\|$, so $F_P(y) - F_P(z) = (1-\varepsilon)(F_Q(y) - F_Q(z)) + \varepsilon\|z-y\|$. But $F_Q(y) - F_Q(z) \geq -\|z-y\|$, so $F_P(y) - F_P(z) \geq (2\varepsilon - 1)\|z-y\|$. Therefore, if $\varepsilon > 1/2$ we have $F_P(y) > F_P(z)$ for all $y \neq z$, implying that $M(P) = z$ and that the median is unique. If $\varepsilon = 1/2$, the weaker inequality $F_P(y) \geq F_P(z)$ holds for all y, so z is still one of the minimizers of F_P but may not be unique.

(ii) If $B_M(\varepsilon) = \infty$, there is a sequence of probabilities $\{Q_n\}$ such that $P_n = (1-\varepsilon)P_0 + \varepsilon Q_n$ and $\|M(P_n)\| \to \infty$. Since $F_{P_n}(y) = (1-\varepsilon)F_{P_0}(y) + \varepsilon F_{Q_n}(y)$, with $F_{Q_n}(y) \geq -\|y\|$ and $F_{P_0}(y)/\|y\| \to 1$ as $\|y\| \to \infty$, it turns out that

$$\liminf F_{P_n}(M(P_n))/\|M(P_n)\| \geq (1 - 2\varepsilon).$$

But $F_{P_n}(M(P_n)) \leq F_{P_n}(0) = 0$, implying that $\varepsilon \geq 1/2$. Therefore $\varepsilon_M^* \geq 1/2$, and since ε_M^* cannot be greater than $1/2$ for a translation equivariant estimator, it follows that $\varepsilon_M^* = 1/2$.

(iii) If $B_M(\varepsilon) \not\to 0$ as $\varepsilon \to 0$, there would be a decreasing sequence $\varepsilon_n \to 0$ with $P_n \in \mathcal{N}_{\varepsilon_n}(P_0)$ such that $\|M(P_n) - \mu\|$ stays bounded away from zero. Since $\|M(P_n) - \mu\|$ is bounded for $\varepsilon_n < 1/2$ and $L^2(T)$ is complete, there would be a subsequence $M(P_{n(k)})$ such that $M(P_{n(k)}) \to \mu^*$ for some $\mu^* \neq \mu$. Then

$$F_{P_{n(k)}}(M(P_{n(k)})) \to F_{P_0}(\mu^*) > F_{P_0}(\mu)$$

(the inequality is strict by uniqueness of the median) and

$$F_{P_{n(k)}}(M(P_{n(k)})) \leq F_{P_{n(k)}}(\mu) \to F_{P_0}(\mu),$$

which is a contradiction.

PROOF OF THEOREM 5

Before giving the proof, we briefly review some functional analysis concepts; for further details see Luenberger (1969). A functional $G : L^2(T) \to \mathbb{R}$ is said to be Fréchet differentiable at y, with derivative $G'(y)$, if $G'(y) : L^2(T) \to \mathbb{R}$ is a bounded linear operator such that $|G(y+h) - G(y) - G'(y)h| = o(\|h\|)$. By 'bounded' we mean that there exists a constant $C < \infty$ such that $|G'(y)h| \leq C\,\|h\|$ for all $h \in L^2(T)$.

Let $\mathcal{B}(L^2(T), \mathbb{R})$ be the space of all bounded linear operators $L^2(T) \to \mathbb{R}$, and for $H \in \mathcal{B}(L^2(T), \mathbb{R})$ define the norm $\|H\|_{\mathcal{B}} := \sup_{\|f\| \leq 1} |Hf|$. We say that G is twice Fréchet differentiable at y, with second derivative $G''(y)$, if $G''(y) : L^2(T) \to \mathcal{B}(L^2(T), \mathbb{R})$ is a bounded linear operator such that $\|G'(y+h) - G'(y) - G''(y)h\|_{\mathcal{B}} = o(\|h\|)$.

Note that by Riesz's Representation Theorem, $H \in \mathcal{B}(L^2(T), \mathbb{R})$ if and only if there is a $h_H \in L^2(T)$ such that $Hf = \langle h_H, f \rangle$, so we will sometimes identify

H with h_H for notational simplicity; thus, if $K : L^2(T) \to \mathcal{B}(L^2(T), \mathbb{R})$ is an invertible operator we will denote $K^{-1}(H)$ by $K^{-1}h_H$.

The identity operator $\mathfrak{I} : L^2(T) \to \mathcal{B}(L^2(T), \mathbb{R})$ is defined as $(\mathfrak{I}f)g = \langle f, g\rangle$ and the tensor product $h_1 \otimes h_2 : L^2(T) \to \mathcal{B}(L^2(T), \mathbb{R})$ as $((h_1 \otimes h_2)f)g = \langle h_1, f\rangle\langle h_2, g\rangle$; again, for notational simplicity we will say that $\mathfrak{I}f = f$ and $(h_1 \otimes h_2)f = \langle h_1, f\rangle g$.

LEMMA 1 (i) *If $P(\{y\}) > 0$ for all $y \in \mathcal{V} \subset L^2(T)$, then $F_P(y)$ is Fréchet differentiable on $\mathcal{V}$. If, in addition, $E_P(\|X - y\|^{-1}) < \infty$ for all $y \in \mathcal{V}$, then $F_P(y)$ is twice Fréchet differentiable on $\mathcal{V}$. The derivatives are given by*

$$\begin{aligned} F'_P(y) &= -E_P\left(\frac{X - y}{\|X - y\|}\right), \\ F''_P(y) &= E_P\left\{\frac{1}{\|X - y\|}\mathfrak{I} - \frac{(X - y) \otimes (X - y)}{\|X - y\|^3}\right\}. \end{aligned}$$

(ii) *Under the conditions of part* (i)*, if $\mu \in \mathcal{V}$ is the median of P, then $F'_P(\mu) \equiv 0$ and $F''_P(\mu)(f, f) \geq 0$ for all $f \neq 0$ (with strict inequality if the median is unique).*
(iii) *Assuming (8) is the central model, if Z has non-degenerate marginals, no atoms in a neighborhood of 0 and $E(\|Z\|^{-1}) < \infty$, then*

$$F''_{P_0}(\mu) = c\mathfrak{I} - \sum_{k=1}^{p} \lambda_k \xi_k (\phi_k \otimes \phi_k),$$

with c and $\xi_1, \ldots, \xi_p$ as in Theorem 5. Moreover, $F''_{P_0}(\mu)$ is invertible and

$$\{F''_{P_0}(\mu)\}^{-1} = \frac{1}{c}\left\{\mathfrak{I} + \sum_{k=1}^{p} \frac{\lambda_k \xi_k}{(c - \lambda_k \xi_k)}(\phi_k \otimes \phi_k)\right\}.$$

Proof. (i) For $F'_P(y)$ given above we have

$$F_P(y+h) - F_P(y) - F'_P(y)h = \int \left(\|x - y - h\| - \|x - y\| + \frac{\langle x - y, h\rangle}{\|x - y\|}\right) P(dx).$$

A second-order Taylor expansion of the function $f(t) = \sqrt{t}$ about $t_0 = \|x-y\|^2$ (assuming $x \neq y$) gives

$$\begin{aligned}\|x-y-h\| &= \|x-y\| + \frac{1}{2\,\|x-y\|}(\|x-y-h\|^2 - \|x-y\|^2) \\ &\quad -\frac{1}{4\eta^{\frac{3}{2}}}\frac{(\|x-y-h\|^2 - \|x-y\|^2)^2}{2},\end{aligned}$$

with η between $\|x-y-h\|^2$ and $\|x-y\|^2$. Then, for $x \neq y$, we have

$$\|x-y-h\| - \|x-y\| + \frac{\langle x-y, h\rangle}{\|x-y\|} = \frac{\|h\|^2}{2} + -\frac{(\|h\|^2 - 2\langle x-y,h\rangle)^2}{8\eta^{\frac{3}{2}}},$$

which implies that $|F_P(y+h) - F_P(y) - F_P'(y)h| = o(\|h\|)$ if $P(\{y\}) = 0$. So F_P is Fréchet differentiable on $\mathcal{V}$ and the Fréchet derivative is $F_P'(y)$ given above.

The functional $F_P'(y)$ is itself Fréchet differentiable if $E_P\{\|X-y\|^{-1}\} < \infty$ for all $y \in \mathcal{V}$. We have

$$\begin{aligned}F_P'(y+h)g - F_P'(y)g - \{F_P''(y)h\}g &= -E_P\Bigg(\frac{\langle X-y-h, g\rangle}{\|X-y-h\|} - \frac{\langle X-y,g\rangle}{\|X-y\|} \\ &\quad + \frac{\langle h,g\rangle}{\|X-y\|} - \frac{\langle X-y,h\rangle\,\langle X-y,g\rangle}{\|X-y\|^3}\Bigg) \\ &= A + B,\end{aligned}$$

with

$$\begin{aligned}A &= -E_P\Bigg\{\langle X-y,g\rangle\left(\frac{1}{\|X-y-h\|} - \frac{1}{\|X-y\|} - \frac{\langle X-y,h\rangle}{\|X-y\|^3}\right)\Bigg\}, \\ B &= -E_P\Bigg(\frac{\langle h,g\rangle}{\|X-y\|} - \frac{\langle h,g\rangle}{\|X-y-h\|}\Bigg).\end{aligned}$$

A second-order Taylor expansion of $f(t) = 1/\sqrt{t}$ about $t_0 = \|x-y\|^2$ will give the bounds $|A| \leq \|g\|\,o(\|h\|)$ and $|B| \leq \|g\|\,o(\|h\|)$, so $\|F_P'(y+h) - F_P'(y) - F_P''(y)h\|_{\mathcal{B}} = o(\|h\|)$.

(ii) Given $f \in L^2(T)$, let $G(s) = F_P(\mu + sf)$, $s \in \mathbb{R}$. Then $G'(0) = F_P'(\mu)f$ and $G''(0) = F_P''(\mu)(f, f)$. If μ is a median of P, G has a minimum at $s = 0$ and then $G'(0) = 0$ and $G''(0) \geq 0$. If the median is unique, 0 is the unique minimizer of G and then $G''(0) > 0$.

(iii) Since Z has no atoms in a neighborhood of 0, it follows from part (A) that F_{P_0} is twice differentiable. The expression of $F_{P_0}''(\mu)$ under model (8) is derived by direct computation:

$$E_{P_0}\left[\frac{\{X(s)-\mu(s)\}\{X(t)-\mu(t)\}}{\|X-\mu\|^3}\right] = \Phi(s)^T \Lambda^{\frac{1}{2}} \Xi \Lambda^{\frac{1}{2}} \Phi(t),$$

with $\Xi = E\{ZZ^T/(Z^T \Lambda Z)^{\frac{3}{2}}\}$; since Z has symmetric marginals, $\Xi = \mathrm{diag}(\xi_1, \ldots, \xi_p)$.

To prove that $F_{P_0}''(\mu)$ is invertible we have to show that given $h \in L^2(T)$ there is a unique $f \in L^2(T)$ such that $\{F_{P_0}''(\mu)f\}g = \langle h, g\rangle$ for all $g \in L^2(T)$; this f will be $\{F_{P_0}''(\mu)\}^{-1}h$. By direct computation $\{F_{P_0}''(\mu)f\}\phi_k = (c - \lambda_k \xi_k)\langle f, \phi_k\rangle$ for $k = 1, \ldots, p$, and if we extend $\phi_1, \ldots, \phi_p$ to an orthonormal basis of $L^2(T)$, $\{\phi_k\}$, we have $\{F_{P_0}''(\mu)f\}\phi_k = c\langle f, \phi_k\rangle$ for $k \geq p+1$. Therefore,

$$\begin{aligned}(c - \lambda_k \xi_k)\langle f, \phi_k\rangle &= \langle h, \phi_k\rangle, \ k = 1, \ldots, p, \\ c\langle f, \phi_k\rangle &= \langle h, \phi_k\rangle, \ k \geq p+1.\end{aligned}$$

Since $f = \sum_{k=1}^{\infty} \langle f, \phi_k\rangle \phi_k$, it follows that

$$f = \sum_{k=1}^{p} \frac{\langle h, \phi_k\rangle}{(c - \lambda_k \xi_k)} \phi_k + \sum_{k=p+1}^{\infty} \frac{\langle h, \phi_k\rangle}{c} \phi_k,$$

and clearly this f is unique since all its basis coefficients are uniquely determined by h; note that $\xi_k > 0$ for all k because the Z_ks are non-degenerate, and $\sum_{k=1}^{p} \lambda_k \xi_k = c$, so $c - \lambda_k \xi_k > 0$ for all k. Finally, since $\sum_{k=p+1}^{\infty} \langle h, \phi_k\rangle \phi_k =$

$h - \sum_{k=1}^{p} \langle h, \phi_k \rangle \phi_k$, we can write

$$f = \frac{1}{c} h + \sum_{k=1}^{p} \left\{ \frac{1}{(c - \lambda_k \xi_k)} - \frac{1}{c} \right\} \langle h, \phi_k \rangle \phi_k,$$

so $\{F''_{P_0}(\mu)\}^{-1}$ is as stated. □

Proof of Theorem 5. The case $z = \mu$ is trivial, so let us consider the case $z \neq \mu$. Since $M(P_{\varepsilon,z}) \to \mu$ as $\varepsilon \to 0$ by Theorem 4, $M(P_{\varepsilon,z}) \neq z$ for any ε small enough. Therefore $P_{\varepsilon,z}(\{M(P_{\varepsilon,z})\}) = 0$ and $M(P_{\varepsilon,z})$ satisfies

$$E_{P_{\varepsilon,z}} \left\{ \frac{\langle z - M(P_{\varepsilon,z}), g \rangle}{\|X - M(P_{\varepsilon,z})\|} \right\} = 0, \text{ for all } g \in L^2(T),$$

or equivalently,

$$(1 - \varepsilon) F'_{P_0}(M(P_{\varepsilon,z}))g - \varepsilon \frac{\langle z - M(P_{\varepsilon,z}), g \rangle}{\|z - M(P_{\varepsilon,z})\|} = 0, \text{ for all } g \in L^2(T).$$

By Lemma 1, the functional $H : [0, 1] \times L^2(T) \to \mathcal{B}(L^2(T), \mathbb{R})$ given by

$$H(\varepsilon, y) = (1 - \varepsilon) F'_{P_0}(y) - \varepsilon \langle \frac{(z - y)}{\|z - y\|}, \cdot \rangle$$

is differentiable at $(\varepsilon, y) = (0, \mu)$ and $D_y H(0, \mu) = F''_{P_0}(\mu)$ is invertible; here $D_y H$ denotes the differential of H with respect to y. The Implicit Function Theorem (Dieudonne, 1969, p. 270) implies that $M(P_{\varepsilon,z})$ is differentiable at $\varepsilon = 0$, so $IF_M(z)$ exists. To find $IF_M(z)$ explicitly, note that

$$\frac{(1 - \varepsilon)}{\varepsilon} \{F'_{P_0}(M(P_{\varepsilon,z}))g - F'_{P_0}(\mu)g\} = \frac{\langle z - M(P_{\varepsilon,z}), g \rangle}{\|z - M(P_{\varepsilon,z})\|}, \text{ for all } g \in L^2(T),$$

and taking $\varepsilon \to 0$ we get

$$\{F''_{P_0}(\mu) IF_M(z)\} g = \frac{\langle z - \mu, g \rangle}{\|z - \mu\|} \text{ for all } g \in L^2(T).$$

Therefore $IF_M(z) = \{F''_{P_0}(\mu)\}^{-1}\{(z-\mu)/\|z-\mu\|\}$.

To find γ^*_M, note that

$$IF_M(z) = \sum_{k=1}^{p} \frac{\langle h, \phi_k\rangle}{(c-\lambda_k\xi_k)}\,\phi_k + \sum_{k=p+1}^{\infty} \frac{\langle h, \phi_k\rangle}{c}\,\phi_k$$

with $h = (z-\mu)/\|z-\mu\|$. So

$$\begin{aligned} \|IF_M(z)\|^2 &\leq \frac{1}{\min_{1\leq k\leq p}(c-\lambda_k\xi_k)^2} \sum_{k=1}^{\infty} |\langle h, \phi_k\rangle|^2 \\ &= \frac{1}{\min_{1\leq k\leq p}(c-\lambda_k\xi_k)^2}\, \|h\|^2, \end{aligned}$$

and $\|h\| = 1$. The upper bound is attained at $h = \phi_{k_0}$ with k_0 the minimizer of $c - \lambda_k\xi_k$.

PROOF OF THEOREM 6

It is a straightforward application of Theorem 19.26 of van der Vaart (1998). In the notation of that book, $\tilde{\mu}_n$ is the solution of $E_{P_n}\{\psi_{\tilde{\mu}_n,h}(X)\} = 0$ for all $h \in L^2(T)$, with $\psi_{\theta,h}(x) = \langle x-\theta, h\rangle/\|x-\theta\|$, $\psi_\theta(x) = (x-\theta)/\|x-\theta\|$ and $\Theta = \{\theta = \mu + f : f \in \text{span}\{\phi_1,\ldots,\phi_p\} \text{ and } \|f\| \leq 1\}$. Note that $P_0(\tilde{\mu}_n \in \Theta) \to 1$ as $n \to \infty$ because $\tilde{\mu}_n$ is consistent and, as pointed out by Kemperman (1987, p. 227), in a Hilbert space the median is contained in any closed convex set that supports P_0. Since Θ is compact in the usual L^2 norm, the family of functions $\{\psi_{\theta,h}(x) : \|\theta-\mu\| < \delta,\, h \in L^2(T)\}$ is P_0-Donsker, $E_{P_0}\{\psi_\theta(X)\} = F'_{P_0}(\theta)$ is Fréchet differentiable at $\theta = \mu$ with derivative $F''_{P_0}(\mu)$ that has a continuous inverse, and $\sup_{\|h\|\leq 1}\|E_{P_0}\{\psi_{\theta,h}(X) - \psi_{\mu,h}(X)\}^2\| \to 0$ as $\theta \to \mu$. Then $F''_{P_0}(\mu)n^{\frac{1}{2}}(\tilde{\mu}_n - \mu) = -\mathbb{G}_n\psi_\mu + o_{P_0}(1)$, where $\mathbb{G}_n$ is the empirical process $n^{\frac{1}{2}}(P_n - P_0)$. This implies that $n^{\frac{1}{2}}(\tilde{\mu}_n - \mu)$ is asymptotically Gaussian with mean zero and covariance functional $E_{P_0}\{IF_M(X) \otimes IF_M(X)\}$, since $IF_M(X) = \{F''_{P_0}(\mu)\}^{-1}\psi_\mu(X)$.

The explicit expression of the covariance kernel $\rho_M(s,t)$ is relatively easy to derive. Since $\{F''_{P_0}(\mu)\}^{-1}$ is self-adjoint, for any $g, h \in L^2(T)$ we obtain

$$
\begin{aligned}
& E_{P_0}\{IF_M(X) \otimes IF_M(X)\}(g,h) \\
= \; & E_{P_0}\left\{\frac{X-\mu}{\|X-\mu\|} \otimes \frac{X-\mu}{\|X-\mu\|}\right\}(\{F''_{P_0}(\mu)\}^{-1}g, \{F''_{P_0}(\mu)\}^{-1}h) \\
= \; & \{\sum_{k=1}^{p} \tilde{\lambda}_k(\phi_k \otimes \phi_k)\}(\{F''_{P_0}(\mu)\}^{-1}g, \{F''_{P_0}(\mu)\}^{-1}h) \\
= \; & \left[\sum_{k=1}^{p} \tilde{\lambda}_k \frac{1}{c^2}\left\{1 + \frac{\lambda_k \xi_k}{(c - \lambda_k \xi_k)}\right\}^2 (\phi_k \otimes \phi_k)\right](g,h) \\
= \; & \left\{\sum_{k=1}^{p} \frac{\tilde{\lambda}_k}{(c - \lambda_k \xi_k)^2}(\phi_k \otimes \phi_k)\right\}(g,h),
\end{aligned}
$$

so the covariance kernel $\rho_M(s,t)$ is almost everywhere as stated in the Theorem.

PROOF OF THEOREM 7

Let $P_{\varepsilon,z} = (1-\varepsilon)P_0 + \varepsilon(0.5\delta_{\mu+z} + 0.5\delta_{\mu-z})$. Since the contaminating distribution is symmetric about μ, we have $M(P_{\varepsilon,z}) = \mu$ for any z. Then the covariance function of $\{X - M(P_{\varepsilon,z})\}/\|X - M(P_{\varepsilon,z})\|$ when $X \sim P_{\varepsilon,z}$ is

$$
\begin{aligned}
\tilde{\rho}_{\varepsilon,z}(s,t) &= (1-\varepsilon)\tilde{\rho}(s,t) + \varepsilon \frac{z(s)z(t)}{\|z\|^2} \\
&= (1-\varepsilon)\sum_{j=1}^{p} \tilde{\lambda}_j \phi_j(s)\phi_j(t) + \varepsilon \frac{z(s)z(t)}{\|z\|^2}.
\end{aligned}
$$

Suppose $k \geq 2$. Taking $z = \phi_k$ we see that the eigenfunctions of $\tilde{\rho}_{\varepsilon,z}$ are $\phi_1, \ldots .\phi_p$ with eigenvalues $(1-\varepsilon)\tilde{\lambda}_j$ for $j \neq k$ and $(1-\varepsilon)\tilde{\lambda}_k + \varepsilon$ for ϕ_k. Therefore, if $(1-\varepsilon)\tilde{\lambda}_k + \varepsilon \geq (1-\varepsilon)\tilde{\lambda}_{k-1}$, $\Phi_k(P_{\varepsilon,z}) = \phi_{k-1}$ and there is breakdown of the kth eigenfunction. Since $(1-\varepsilon)\tilde{\lambda}_k + \varepsilon \geq (1-\varepsilon)\tilde{\lambda}_{k-1}$ if and only if $\varepsilon \geq (\tilde{\lambda}_{k-1} - \tilde{\lambda}_k)/(1 + \tilde{\lambda}_{k-1} - \tilde{\lambda}_k)$, we have $\varepsilon^*_{\Phi_k} \leq (\tilde{\lambda}_{k-1} - \tilde{\lambda}_k)/(1 + \tilde{\lambda}_{k-1} - \tilde{\lambda}_k)$ for $k \geq 2$. Similarly, for $k \leq p-1$ take $z = \phi_{k+1}$ and observe that $\Phi_k(P_{\varepsilon,z}) = \phi_{k+1}$

if $(1-\varepsilon)\tilde{\lambda}_{k+1}+\varepsilon \geq (1-\varepsilon)\tilde{\lambda}_k$, which occurs if and only if $\varepsilon \geq (\tilde{\lambda}_k - \tilde{\lambda}_{k+1})/(1+\tilde{\lambda}_k - \tilde{\lambda}_{k+1})$; this gives the upper bound $\varepsilon^*_{\Phi_k} \leq (\tilde{\lambda}_k - \tilde{\lambda}_{k+1})/(1+\tilde{\lambda}_k - \tilde{\lambda}_{k+1})$ for $k \leq p-1$.

PROOF OF THEOREM 8

Let $P_{\varepsilon,z} = (1-\varepsilon)P_0 + \varepsilon\delta_z$. Then $\Phi_k(P_{\varepsilon,z})$ is, by definition, the kth eigenfunction of

$$\begin{aligned}\tilde{\rho}_{\varepsilon,z}(s,t) &= E_{P_{\varepsilon,z}}\left[\frac{\{X(s)-M(P_{\varepsilon,z})(s)\}\{X(t)-M(P_{\varepsilon,z})(t)\}}{\|X-M(P_{\varepsilon,z})\|^2}\right] \qquad (1)\\ &= (1-\varepsilon)E_{P_0}\left[\frac{\{X(s)-M(P_{\varepsilon,z})(s)\}\{X(t)-M(P_{\varepsilon,z})(t)\}}{\|X-M(P_{\varepsilon,z})\|^2}\right]\\ &\quad +\varepsilon\frac{\{z(s)-M(P_{\varepsilon,z})(s)\}\{z(t)-M(P_{\varepsilon,z})(t)\}}{\|z-M(P_{\varepsilon,z})\|^2}.\end{aligned}$$

That is,

$$\int \tilde{\rho}_{\varepsilon,z}(s,t)\Phi_k(P_{\varepsilon,z})(s)ds = \tilde{\Lambda}_k(P_{\varepsilon,z})\Phi_k(P_{\varepsilon,z})(t),\ t\in T. \qquad (2)$$

Note that (1) is valid only if $M(P_{\varepsilon,z}) \neq z$, which is the case for any ε small enough as long as $z \neq \mu$. When $z=\mu$, $M(P_{\varepsilon,z})=\mu$ for all ε and then $\tilde{\rho}_{\varepsilon,z}(s,t) = (1-\varepsilon)\tilde{\rho}_0(s,t)$, $\Phi_k(P_{\varepsilon,z})=\phi_k$ and $\tilde{\Lambda}_k(P_{\varepsilon,z}) = (1-\varepsilon)\tilde{\lambda}_k$. Therefore, $IF_{\Phi_k}(\mu)=0$ and $IF_{\tilde{\Lambda}_k}(\mu) = -\tilde{\lambda}_k$.

Let us now consider the case $z\neq\mu$. Let $H:[0,1]\times L^2(T)\times(0,\infty)\times L^2(T)\to L^2(T)\times\mathbb{R}$ be given by

$$\begin{aligned}H_1(\varepsilon,y,\lambda,\phi) &= (1-\varepsilon)E_{P_0}\left\{\frac{\langle X-y,\phi\rangle(X-y)}{\|X-y\|^2}\right\}+\varepsilon\frac{\langle z-y,\phi\rangle(z-y)}{\|z-y\|^2}\\ &\quad -\lambda\phi,\\ H_2(\varepsilon,y,\lambda,\phi) &= \|\phi\|^2-1.\end{aligned}$$

Then $\tilde{\Lambda}_k(P_{\varepsilon,z})$ and $\Phi_k(P_{\varepsilon,z})$ are implicitly defined by the equation $H(\varepsilon, M(P_{\varepsilon,z}),$

$\tilde{\Lambda}_k(P_{\varepsilon,z}), \Phi_k(P_{\varepsilon,z})) = 0$. The partial derivatives of H_1 and H_2 with respect to λ and ϕ are: $D_\lambda H_1 = -\phi$,

$$D_\phi H_1 = (1-\varepsilon) E_{P_0}\left\{ \frac{(X-y)\otimes(X-y)}{\|X-y\|^2} \right\} + \varepsilon \frac{(z-y)\otimes(z-y)}{\|z-y\|^2} - \lambda \mathfrak{I},$$

$D_\lambda H_2 = 0$ and $D_\phi H_2 = 2\phi$. The operator $[D_\lambda H_1; D_\phi H_1; D_\lambda H_2; D_\phi H_2]$ at $(\varepsilon, y, \lambda, \phi) = (0, \mu, \lambda_k, \phi_k)$ is invertible, since we are assuming the kth eigenvalue has multiplicity one. Then the Implicit Function Theorem guarantees that $\Phi_k(P_{\varepsilon,z})$ and $\tilde{\Lambda}_k(P_{\varepsilon,z})$ are well-defined in a neighborhood of $(0, \mu, \lambda_k, \phi_k)$ and that they are differentiable at $\varepsilon = 0$. So we can differentiate with respect to ε on both sides of (2) and get

$$\begin{aligned} & \int \frac{\partial}{\partial \varepsilon} \tilde{\rho}_{\varepsilon,z}(s,t)\Big|_{\varepsilon=0} \phi_k(s) ds + \int \tilde{\rho}_0(s,t) IF_{\Phi_k}(z)(s) ds \\ = \;& IF_{\tilde{\Lambda}_k}(z)\phi_k(t) + \tilde{\lambda}_k IF_{\Phi_k}(z)(t) \text{ for all } t \in T. \end{aligned} \tag{3}$$

Since

$$\begin{aligned} \frac{\partial}{\partial \varepsilon} \tilde{\rho}_{\varepsilon,z}(s,t)\Big|_{\varepsilon=0} = \;& -\tilde{\rho}_0(s,t) - IF_M(z)(s) E_{P_0}\left\{ \frac{X(t)-\mu(t)}{\|X-\mu\|^2} \right\} \\ & - E_{P_0}\left\{ \frac{X(s)-\mu(s)}{\|X-\mu\|^2} \right\} IF_M(z)(t) \\ & + 2E_{P_0}\left[\frac{\{X(s)-\mu(s)\}\{X(t)-\mu(t)\}\langle X-\mu, IF_M(z)\rangle}{\|X-\mu\|^4} \right] \\ & + \frac{\{z(s)-\mu(s)\}\{z(t)-\mu(t)\}}{\|z-\mu\|^2}, \end{aligned}$$

and all the above expectations vanish by the symmetry of $X - \mu$, we have

$$\int \frac{\partial}{\partial \varepsilon} \tilde{\rho}_{\varepsilon,z}(s,t)\Big|_{\varepsilon=0} \phi_k(s) ds = -\tilde{\lambda}_k \phi_k(t) + \frac{\{z(t)-\mu(t)\}\langle z-\mu, \phi_k\rangle}{\|z-\mu\|^2}.$$

Substituting in (3) we obtain

$$\begin{aligned} & -\tilde{\lambda}_k\phi_k(t) + \frac{\{z(t)-\mu(t)\}\left<z-\mu,\phi_k\right>}{\|z-\mu\|^2} + \int \tilde{\rho}_0(s,t) IF_{\Phi_k}(z)(s)ds \quad (4) \\ = \ & IF_{\tilde{\Lambda}_k}(z)\phi_k(t) + \tilde{\lambda}_k IF_{\Phi_k}(z)(t) \text{ for all } t \in T. \end{aligned}$$

Taking inner products with ϕ_k on both sides of (4), and using that $\langle \phi_k, IF_{\Phi_k}(z)\rangle = 0$ because $\|\Phi_k(P_{\varepsilon,z})\|^2 = 1$ for all ε, we have

$$-\tilde{\lambda}_k + \frac{\langle z-\mu,\phi_k\rangle^2}{\|z-\mu\|^2} = IF_{\tilde{\Lambda}_k}(z).$$

Taking inner products with ϕ_j, $j = 1,\ldots,p$, $j \neq k$, on both sides of (4) we get

$$\frac{\langle z-\mu,\phi_j\rangle\langle z-\mu,\phi_k\rangle}{\|z-\mu\|^2} + \tilde{\lambda}_j\langle \phi_j, IF_{\Phi_k}(z)\rangle = \tilde{\lambda}_k\langle \phi_j, IF_{\Phi_k}(z)\rangle.$$

If $\phi_1,\ldots,\phi_p$ is extended to an orthonormal basis of $L^2(T)$, $\{\phi_j\}$, by taking inner products with ϕ_j, $j > p$, on both sides of (4), we get

$$\frac{\langle z-\mu,\phi_j\rangle\langle z-\mu,\phi_k\rangle}{\|z-\mu\|^2} = \tilde{\lambda}_k\langle \phi_j, IF_{\Phi_k}(z)\rangle.$$

Therefore

$$\begin{aligned} IF_{\Phi_k}(z) &= \sum_{j=1}^{\infty}\langle \phi_j, IF_{\Phi_k}(z)\rangle\phi_j \\ &= \sum_{\substack{j=1\\ j\neq k}}^{p}\frac{\zeta_j(z)\zeta_k(z)}{(\tilde{\lambda}_k-\tilde{\lambda}_j)}\phi_j + \sum_{j=p+1}^{\infty}\frac{\zeta_j(z)\zeta_k(z)}{\tilde{\lambda}_k}\phi_j \\ &= \frac{\zeta_k(z)}{\tilde{\lambda}_k}\Bigg\{\sum_{\substack{j=1\\ j\neq k}}^{p}\frac{\tilde{\lambda}_k\zeta_j(z)}{(\tilde{\lambda}_k-\tilde{\lambda}_j)}\phi_j + \frac{z-\mu}{\|z-\mu\|} - \zeta_k(z)\phi_k - \sum_{\substack{j=1\\ j\neq k}}^{p}\zeta_j(z)\phi_j\Bigg\}, \end{aligned}$$

as stated. Moreover,

$$\begin{aligned} \|IF_{\Phi_k}(z)\|^2 &= \zeta_k(z)^2 \left\{ \sum_{\substack{j=1 \\ j \neq k}}^{p} \frac{\zeta_j(z)^2}{(\tilde{\lambda}_k - \tilde{\lambda}_j)^2} + \sum_{j=p+1}^{\infty} \frac{\zeta_j(z)^2}{\tilde{\lambda}_k^2} \right\} \\ &\leq \frac{1}{c_k^2} \zeta_k(z)^2 \{1 - \zeta_k(z)^2\}, \end{aligned}$$

since $\sum_{j=1}^{\infty} \zeta_j(z)^2 = 1$. Therefore $\|IF_{\Phi_k}(z)\|^2 \leq 1/(4c_k^2)$ for any $z \in L^2(T)$ and the upper bound is attained at $z^* = \mu + (\phi_k + \phi_{j^*})/\sqrt{2}$, where j^* is such that $|\tilde{\lambda}_k - \tilde{\lambda}_{j^*}| = c_k$ if $c_k = \min\{|\tilde{\lambda}_k - \tilde{\lambda}_j| : j = 1, \ldots, p,\ j \neq k\}$, and j^* can be any j larger than or equal to $p + 1$ if $c_k = \tilde{\lambda}_k$.

REFERENCES

DIEUDONNE, J. (1969). *Foundations of Modern Analysis.* New York: Academic Press.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 221–233. University of California Press.

LUENBERGER, D. G. (1969). *Optimization by Vector Space Methods*. New York: John Wiley.

MASSE, J.-C. (1997). Nonparametric maximum likelihood estimation in a non locally compact setting. *Metrika* **46** 123–145.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.