Supplementary Material for *Deconvolution When Classifying Noisy Data Involving Transformations*

Raymond Carroll
Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, USA
carroll@stat.tamu.edu

Aurore Delaigle
Department of Mathematics and Statistics, University of Melbourne, VIC, 3010, Australia
A.Delaigle@ms.unimelb.edu.au

Peter Hall
Department of Mathematics and Statistics, University of Melbourne, VIC, 3010, Australia
halpstat@ms.unimelb.edu.au

A Supplement to Numerical Implementation

A.1 Breaking ties for the centroid classifier

In cases where $\widehat{e}(\theta)$ achieves its minimum at several values θ, we break the ties by choosing θ to minimize

$$-\frac{\pi_1}{n_1 \cdot M(\theta)} \sum_{i=1}^{n_1} S_{\theta;-i1}(Y_{i1})\, I\{S_{\theta;-i1}(Y_{i1}) \le 0\}$$
$$+\frac{1-\pi_1}{n_2 \cdot M(\theta)} \sum_{i=1}^{n_2} S_{\theta;-i2}(Y_{i2})\, I\{S_{\theta;-i2}(Y_{i2}) > 0\}$$

over θ for which $\widehat{e}(\theta)$ is tied at the least value it takes. Here, $M(\theta)$ is a normalizing factor defined by $M(\theta) = \max_{j=1,2}\ n_j^{-1} \sum_{i=1}^{n_j} \sum_{r\in\mathcal{D}} \left|Z_{ij;\theta}(r) - \bar{Z}_j^{(-i)}(r;\theta)\right|^2$.

A.2 Other choices of spread function ω_{Q_θ}

Alternatives to the two-parameter family $\omega_{p_0;\,\theta}$ introduced in Section 2.3 include taking ω_{Q_θ} to have relatively narrow support, and choosing the weights $\omega_{Q_\theta}(r)$ to be small in number, for example to equal θ_j if $c_{j-1} < \|r\| \le c_j$ for $1 \le j \le q$ and $r \in \mathbb{Z}^d$, and to equal 0 otherwise. Here, $\|r\|^2 = \sum_{1\le j\le d} r_j^2$, $0 = c_0 < c_1 < \ldots < c_q$ are constants, $\theta_1 \ge \ldots \ge \theta_q > 0$, and the weights would usually be normalized so that $\sum_{r\in\mathbb{Z}^d} \omega_{Q_\theta}(r) = 1$.

Remark 1. In some instances we may have in mind a choice of models for Q_θ, potentially with different values, q, of the length of θ. We can choose among models, including among

Table 1: Values of σ, θ_R and signal to noise ratio for models (a) and (b)

Model (a)									
ρ_R	0.75	0.5	0.25	0.75	0.5	0.25	0.75	0.5	0.25
ℓ_R	3	3	3	2	2	2	1	1	1
σ	5	1	1	2	1	1	1	1	.8
SNR ($k_{\mathcal{I}} = 1$)	0.5682	1.8097	1.2231	1.2600	1.6007	1.0687	1.9844	1.2450	1.0484
SNR ($k_{\mathcal{I}} = 2$)	0.4187	1.3304	0.8329	0.9278	1.1539	0.6645	1.4082	0.7867	0.5793
Model (b)									
ρ_R	0.75	0.5	0.25	0.75	0.5	0.25	0.75	0.5	0.25
ℓ_R	3	3	3	2	2	2	1	1	1
σ	1	0.5	0.5	1	0.5	0.25	0.5	0.5	0.25
SNR ($k_{\mathcal{I}} = 1$)	1.6122	2.0540	1.3882	1.4300	1.8167	2.4258	2.2523	1.4130	1.9039
SNR ($k_{\mathcal{I}} = 2$)	1.9603	2.4915	1.5599	1.7376	2.1610	2.4888	2.6372	1.4733	1.7359

values of q, by using crossvalidation to assess the impact that this choice has on classification error, or by using Monte Carlo simulation from models that seem to mimic the data well.

A.3 Complements to numerical section

A.3.1 Signal to noise ratio

For each example, we calculated a signal to noise ratio SNR, which we defined as

$$\text{SNR} = \max_{r \in \mathcal{D}} \frac{|T\{\mu_2(r) - \mu_1(r)\}|}{\sqrt{\text{var}\{R\,\xi(r)\}}}. \tag{A.1}$$

The values of SNR are shown in Tables 1 to 3 for each example considered in our numerical work.

A.3.2 Additional figures for Section 4.3.3

Here we show additional results obtained when applying to inversion method to the centroid classifier when the errors errors $R_r\,\xi_{ij}(r)$ were non stationary. As in Section 4.3.3, we inverted the data through $Q^{-1}_{\widehat{\theta}_{CV}}$. Figures 1 and 2 show boxplots of the percentage of missclassified curves for the centroid classifier constructed from the data $Q^{-1}_{\widehat{\theta}_{CV}} Y_{ij}$, Y_{ij} and $T^{-1}Y_{ij}$, where Y_{ij} was generated as in (4.5), with, μ_j from model (a) and model (b), respectively, R_r as in (4.6) and $\alpha = 10$.

Table 2: Values of σ, θ_R and signal to noise ratio for models (c) and (d)

Model (c)						
ρ_R	0.85	0.5	0.85	0.5	0.85	0.5
ℓ_R	3	3	2	2	1	1
σ	5	1	5	1	2	.5
SNR ($k_{\mathcal{I}} = 1$)	1.5615	1.7838	1.2291	1.3955	1.9155	1.6884
SNR ($k_{\mathcal{I}} = 2$)	1.3669	1.5540	1.0758	1.1690	1.6344	1.0868
Model (d)						
ρ_R	0.85	0.5	0.85	0.5	0.85	0.5
ℓ_R	3	3	2	2	1	1
σ	2	1	2	0.5	1	.5
SNR ($k_{\mathcal{I}} = 1$)	0.6401	0.2925	0.5039	0.4577	0.6282	0.2769
SNR ($k_{\mathcal{I}} = 2$)	0.6101	0.2775	0.4802	0.4174	0.5836	0.1940

Table 3: Values of σ, θ_M and signal to noise ratio for models (c) and (d)

Model (c)			
θ_M	30	20	10
σ	5	2.5	1
SNR ($k_{\mathcal{I}} = 1$)	1.4531	1.9676	2.5689
SNR ($k_{\mathcal{I}} = 2$)	0.3182	0.4311	0.5645
Model (d)			
θ_M	30	20	10
σ	2.5	1	1
SNR ($k_{\mathcal{I}} = 1$)	0.4766	0.8066	0.4213
SNR ($k_{\mathcal{I}} = 2$)	0.1136	0.1924	0.1008

Figure 3 shows similar results for the bivariate examples (c) and (d), when the data were generated according to (4.5) with R_r as in (4.7) and $\alpha = 4$. These results are similar to those discussed in Section 4.3.3 of the paper.

A.3.3 Logistic classifier based on PLS projection

Let $I_{ij} = j - 1$, where Y_{ij} comes from population Π_j, with $j = 1$ or 2, and let D_{ij} denote one of the four versions of the observations that we use to construct our classifiers (i.e., D_{ij} denotes any of Y_{ij}, $T^{-1}Y_{ij}$, $R^{-1}Y_{ij}$ or $\widehat{Q}^{-1}Y_{ij}$). The PLS logistic classifier assumes that the regression

$$g(x) = E(I_{ij} \mid D_{ij} = x)$$

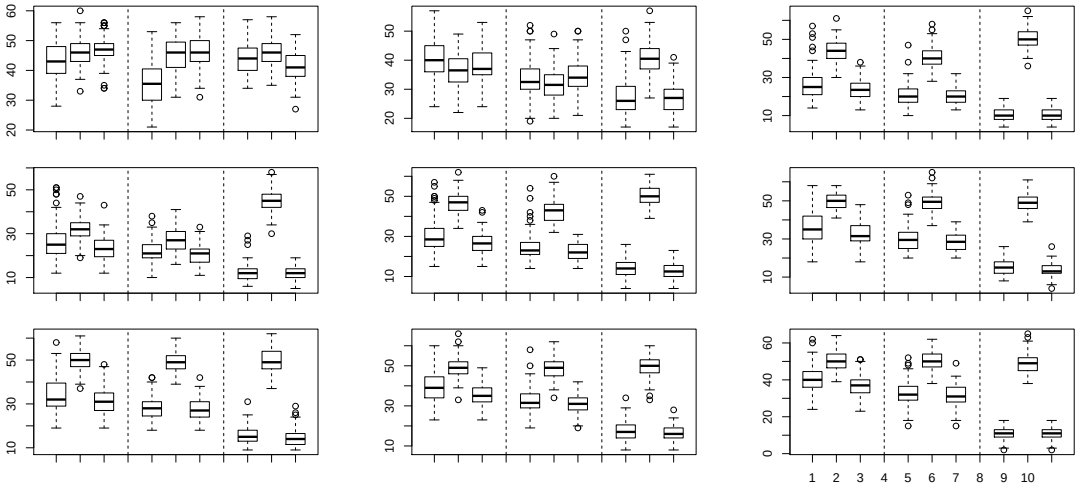

Figure 1: Boxplots of percentage of misclassified observations calculated from 100 simulated samples from model (a) when $\theta_R = (\rho_{R,r}, \ell_R)$, with $\rho_{R,r} = \rho_R + 0.1\cos(r/10)$, $\rho_R = 0.75$, 0.5 and 0.25 in rows 1, 2 and 3, respectively, and $\ell_R = 3$, 2 and 1 in columns 1, 2 and 3, respectively. In each group of 9 boxes, the first three are for $n_1 = n_2 = 10$ and $k_{\mathcal{I}} = 2$, the next three are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 2$, and the last three are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 1$. In each group of three boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), T^{-1} (second box), or untransformed (third box).

follows a logistic model,

$$g(x) = \frac{\exp(\alpha_0 + \alpha_1 \int_{\mathcal{D}} x\beta)}{1 + \exp(\alpha_0 + \alpha_1 \int_{\mathcal{D}} x\beta)}, \tag{A.2}$$

where α_0 and $\alpha_1 \in \mathbb{R}$ and β is a function defined on $\mathcal{D}$. With the PLS approach, β is taken equal to the slope, β, of the linear approximation $\beta_0 + \int_{\mathcal{D}} \beta D_{ij}$ to the regression function $E(I_{ij} \mid D_{ij})$. Then, β is estimated by the estimator $\widehat{\beta}$ obtained by PLS (see Delaigle and Hall (2012a,b) for details about the PLS estimator), and α_0 and α_1 are estimated by $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$, obtained by a least-squares fit of I_{ij} on $\int_{\mathcal{D}} D_{ij}\widehat{\beta}$, using the logistic model (A.2) with β replaced by $\widehat{\beta}$.

The regression function g is then estimated by

$$\widehat{g}(x) = \frac{\exp(\widehat{\alpha}_0 + \widehat{\alpha}_1 \int_{\mathcal{D}} x\widehat{\beta})}{1 + \exp(\widehat{\alpha}_0 + \widehat{\alpha}_1 \int_{\mathcal{D}} x\widehat{\beta})},$$

and the classifier assigns a new data function D to population Π_1 if $\widehat{g}(D) < 0.5$ and to population Π_2 otherwise. The partial least-squares slope estimator $\widehat{\beta}$ depends on a smoothing parameter m, which is the number of PLS basis functions employed to calculated $\widehat{\beta}$. This m is chosen by crossvalidation, together with θ.

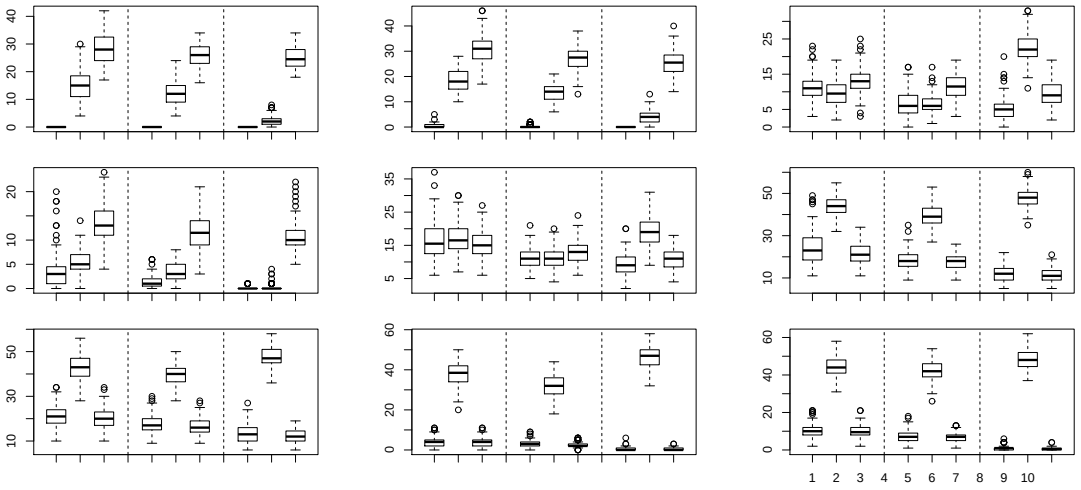

Figure 2: Boxplots of percentage of misclassified observations calculated from 100 simulated samples from model (b) when $\theta_R = (\rho_{R,r}, \ell_R)$, with $\rho_{R,r} = \rho_R + 0.1\cos(r/10)$, $\rho_R = 0.75$, 0.5 and 0.25 in rows 1, 2 and 3, respectively, and $\ell_R = 3$, 2 and 1 in columns 1, 2 and 3, respectively. In each group of 9 boxes, the first three are for $n_1 = n_2 = 10$ and $k_{\mathcal{I}} = 2$, the next three are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 2$, and the last three are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 1$. In each group of three boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), T^{-1} (second box), or untransformed (third box).

A.3.4 Simulation results for unbalanced samples

We also applied the centroid classifier to unbalanced training samples. The results for models (a) and (b), with $n_1 = 10$, $n_2 = 25$ and $k_{\mathcal{I}} = 1$ or $k_{\mathcal{I}} = 2$, are shown in Figures 4 and 5. We can see that the results are similar to those we obtained when $n_1 = n_2$.

A.3.5 Simulation results for other classifiers

We applied the SVM classifier and the logistic PLS classifier to each of the four versions of the data: the untransformed noisy data Y_i^{New}, the data $T^{-1}Y_i^{\text{New}}$, the data $\widehat{Q}^{-1}Y_i^{\text{New}}$ and the data $R^{-1}Y_i^{\text{New}}$. We used the same simulation settings as in Section 4 (same models, same values of θ_R and θ_T, same training and test sample sizes, etc), and as in Section 4 we report the results of our simulations by showing boxplots of the number of observations misclassified by each classifier, for each version of the data. In the case $\widehat{Q}^{-1}Y_i^{\text{New}}$ we chose θ by minimizing the crossvalidation criterion defined in (2.12), taking $\mathcal{C}$ to be the SVM classifier or the logistic PLS classifier. In case of ties for the SVM classifier, we took the smallest value of ρ and θ that minimize crossvalidation. For the logistic classifier, we broke the ties using the method suggested in Delaigle and Hall (2012b).

Figures 6 to 9 show the results for the SVM classifier, for models (a) to (d), training

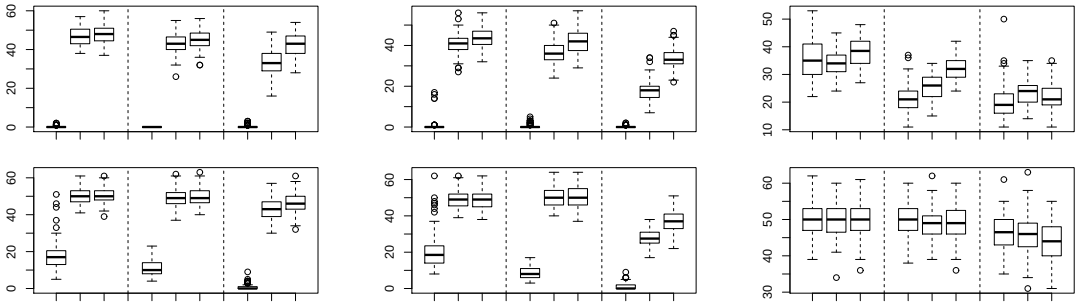

Figure 3: Boxplots of percentage of misclassified observations calculated from 100 simulated samples from models (c) (row 1) and (d) (rows 2) when R_r is of the form at (4.7), with $\theta_{M,r_j} = \theta_M + 2 \cdot [4\cos(r_j/2)]$ and $\theta_M = 30$, 20 and 10 in columns 1, 2 and 3, respectively. In each group of 9 boxes, the first three are for $n_1 = n_2 = 10$ and $k_{\mathcal{I}} = 2$, the next three are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 2$, and the last three are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 1$. In each group of three boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), T^{-1} (second box), or untransformed (third box).

sample sizes $n_1 = n_2 = 10$ or $n_1 = n_2 = 25$, and when the grid $\mathcal{D}$ has edge width $k_{\mathcal{I}} = 1$ or $k_{\mathcal{I}} = 2$. Figures 10 and 11 show the results for the logistic classifier, for models (a) and (b) for training samples of sizes $n_1 = n_2 = 10$ or $n_1 = n_2 = 25$, and when the grid $\mathcal{D}$ has edge width $k_{\mathcal{I}} = 2$.

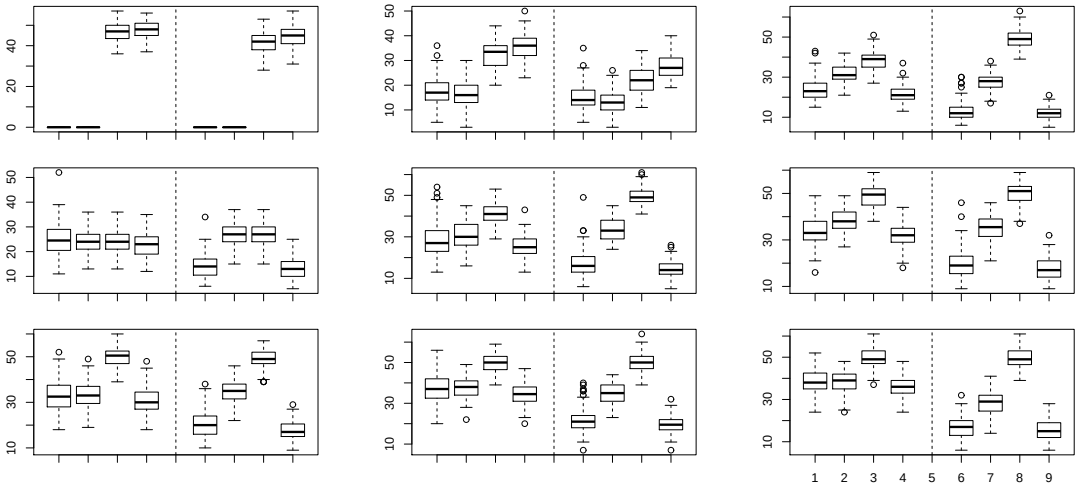

Figure 4: Boxplots of percentage of misclassified observations calculated from 100 simulated samples from model (a) when $\theta_R = (\rho_R, \ell_R)$, with $\rho_R = 0.75$, 0.5 and 0.25 in rows 1, 2 and 3, respectively, and $\ell_R = 3$, 2 and 1 in columns 1, 2 and 3, respectively. In each group of 8 boxes, the first four are for $n_1 = 10$, $n_2 = 25$ and $k_{\mathcal{I}} = 2$, and the next four are for $n_1 = 10$, $n_2 = 25$ and $k_{\mathcal{I}} = 1$. In each group of four boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), R^{-1} (second box), T^{-1} (third box), or untransformed (fourth box).

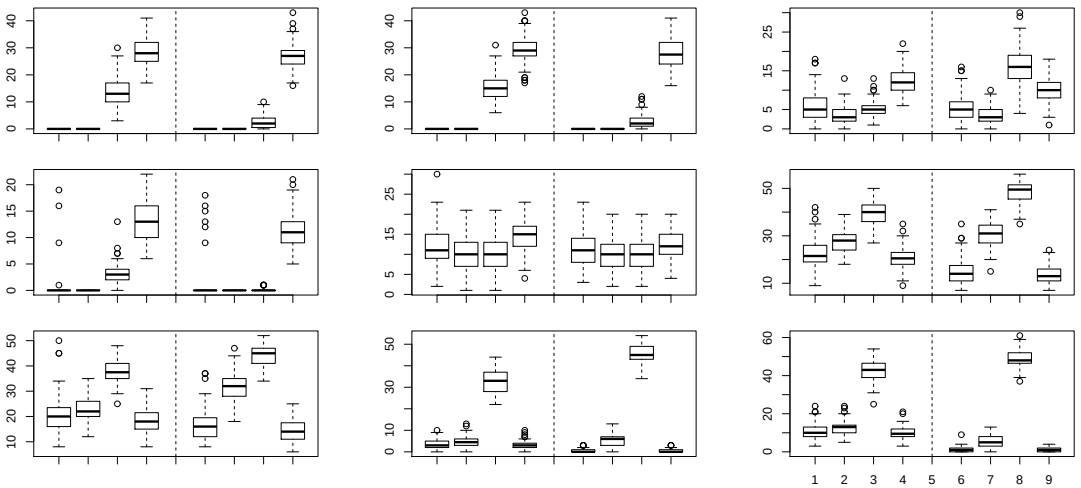

Figure 5: Boxplots of percentage of misclassified observations calculated from 100 simulated samples from model (b) when $\theta_R = (\rho_R, \ell_R)$, with $\rho_R = 0.75$, 0.5 and 0.25 in rows 1, 2 and 3, respectively, and $\ell_R = 3$, 2 and 1 in columns 1, 2 and 3, respectively. In each group of 8 boxes, the first four are for $n_1 = 10$, $n_2 = 25$ and $k_{\mathcal{I}} = 2$, and the next four are for $n_1 = 10$, $n_2 = 25$ and $k_{\mathcal{I}} = 1$. In each group of four boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), R^{-1} (second box), T^{-1} (third box), or untransformed (fourth box).

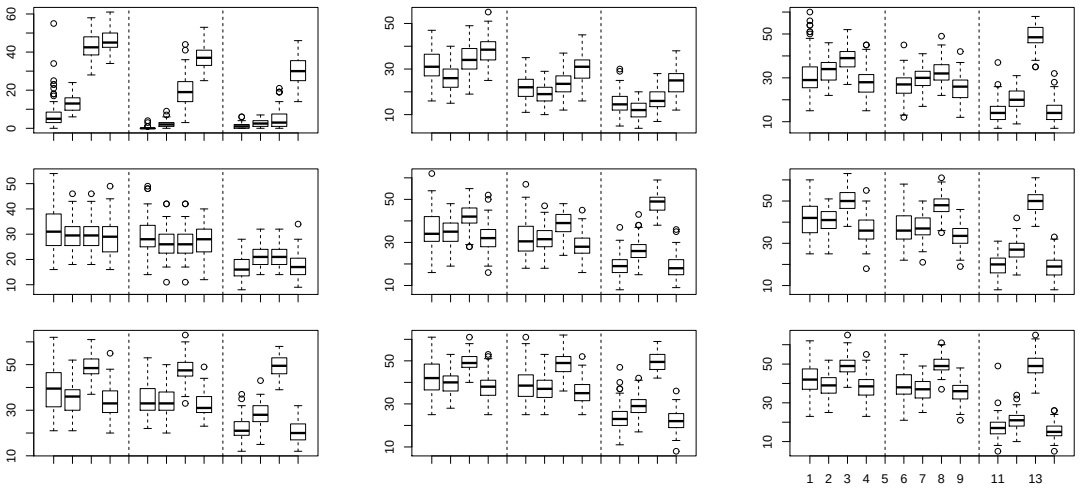

Figure 6: Boxplots of percentage of observations misclassified by the SVM classifier, calculated from 100 simulated samples from model (a) when $\theta_R = (\rho_R, \ell_R)$, with $\rho_R = 0.75$, 0.5 and 0.25 in rows 1, 2 and 3, respectively, and $\ell_R = 3$, 2 and 1 in columns 1, 2 and 3, respectively. In each group of 12 boxes, the first four are for $n_1 = n_2 = 10$ and $k_{\mathcal{I}} = 2$, the next four are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 2$, and the last four are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 1$. In each group of four boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), R^{-1} (second box), T^{-1} (third box), or untransformed (fourth box).

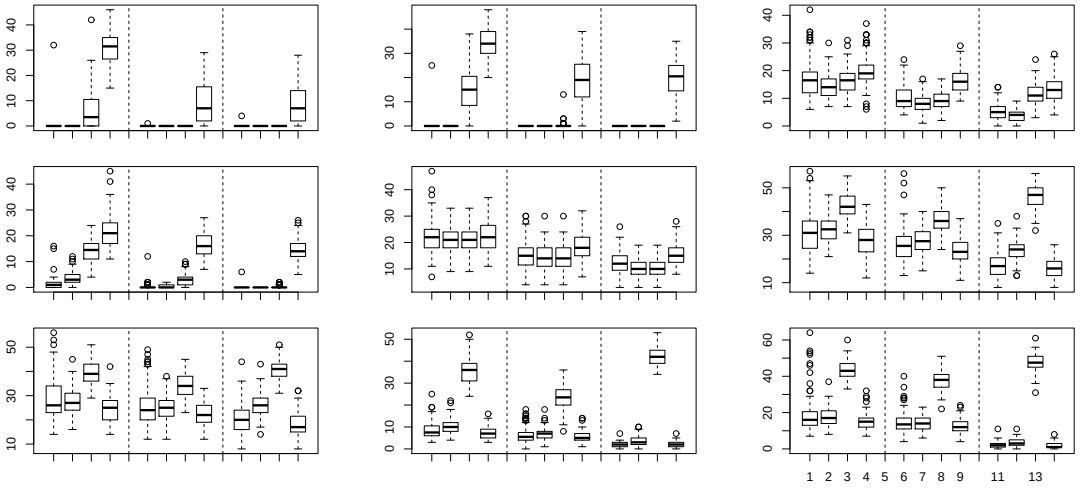

Figure 7: Boxplots of percentage of observations misclassified by the SVM classifier, calculated from 100 simulated samples from model (b) when $\theta_R = (\rho_R, \ell_R)$, with $\rho_R = 0.75$, 0.5 and 0.25 in rows 1, 2 and 3, respectively, and $\ell_R = 3$, 2 and 1 in columns 1, 2 and 3, respectively. In each group of 12 boxes, the first four are for $n_1 = n_2 = 10$ and $k_{\mathcal{I}} = 2$, the next four are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 2$, and the last four are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 1$. In each group of four boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), R^{-1} (second box), T^{-1} (third box), or untransformed (fourth box).

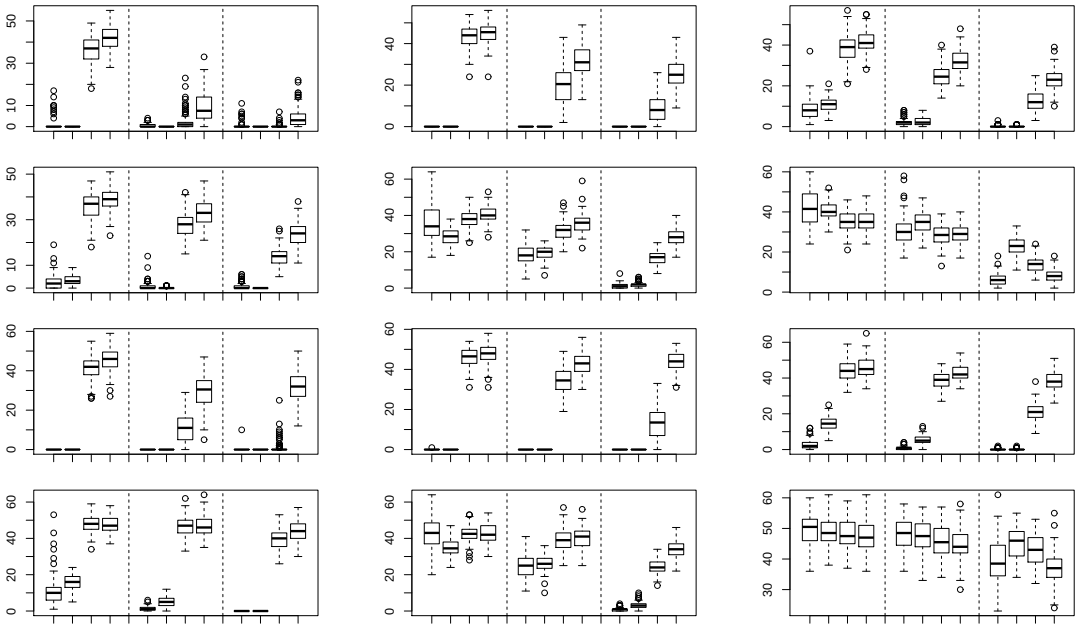

Figure 8: Boxplots of percentage of observations misclassified by the SVM classifier, calculated from 100 simulated samples from models (c) (rows 1 and 2) and (d) (rows 3 and 4) when $\theta_R = (\rho_R, \ell_R)$, with $\rho_R = 0.85$ and 0.5 in rows 1,3 and 2,4 respectively, and $\ell_R = 3$, 2 and 1 in columns 1, 2 and 3, respectively. In each group of 12 boxes, the first four are for $n_1 = n_2 = 10$ and $k_{\mathcal{I}} = 2$, the next four are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 2$, and the last four are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 1$. In each group of four boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), R^{-1} (second box), T^{-1} (third box), or untransformed (fourth box).

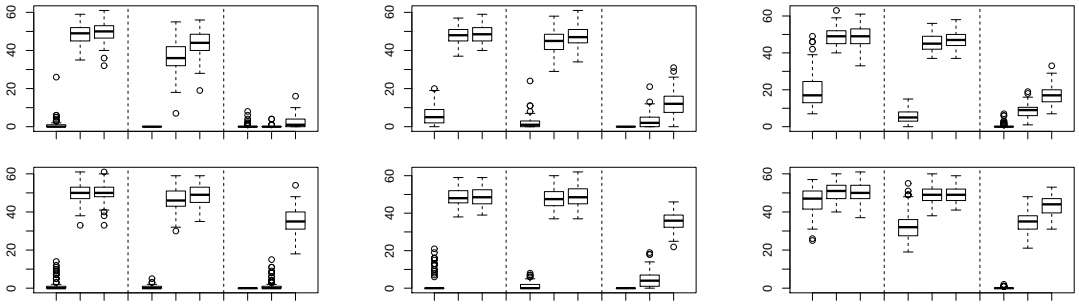

Figure 9: Boxplots of percentage of observations misclassified by the SVM classifier, calculated from 100 simulated samples from models (c) (row 1) and (d) (rows 2) when R is of the form at (4.3), with $\theta_M = 30$, 20 and 10 in columns 1, 2 and 3, respectively. In each group of 9 boxes, the first three are for $n_1 = n_2 = 10$ and $k_{\mathcal{I}} = 2$, the next three are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 2$, and the last three are for $n_1 = n_2 = 25$ and $k_{\mathcal{I}} = 1$. In each group of three boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), T^{-1} (second box), or untransformed (third box).

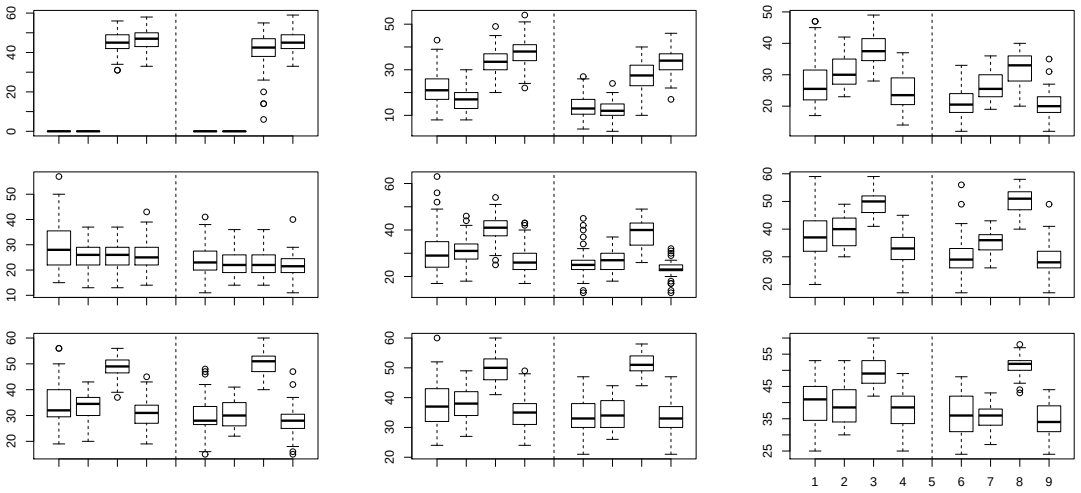

Figure 10: Boxplots of percentage of observations misclassified by the logistic classifier, calculated from 100 simulated samples from model (a) when $\theta_R = (\rho_R, \ell_R)$, with $\rho_R = 0.75$, 0.5 and 0.25 in rows 1, 2 and 3, respectively, and $\ell_R = 3$, 2 and 1 in columns 1, 2 and 3, respectively. In each group of 8 boxes, the first four are for $n_1 = n_2 = 10$ and $k_{\mathcal{I}} = 2$, and the next four are for $n_1 = n_2 = 25$. In each group of four boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), R^{-1} (second box), T^{-1} (third box), or untransformed (fourth box).

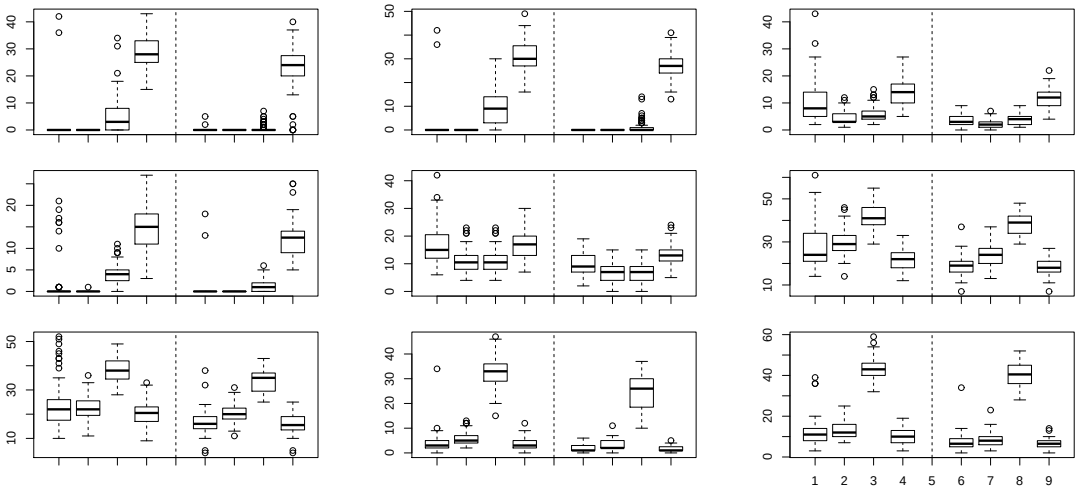

Figure 11: Boxplots of percentage of observations misclassified by the logistic classifier, calculated from 100 simulated samples from model (b) when $\theta_R = (\rho_R, \ell_R)$, with $\rho_R = 0.75$, 0.5 and 0.25 in rows 1, 2 and 3, respectively, and $\ell_R = 3$, 2 and 1 in columns 1, 2 and 3, respectively. In each group of 8 boxes, the first four are for $n_1 = n_2 = 10$ and $k_{\mathcal{I}} = 2$, and the next four are for $n_1 = n_2 = 25$. In each group of four boxes, the data are transformed by $\widehat{Q}^{-1}$ (first box), R^{-1} (second box), T^{-1} (third box), or untransformed (fourth box).

B Technical Arguments

B.1 Regularity conditions

We define

$$q = \int_{\mathcal{T}} |\phi_{\omega_R}|^4 \sup_{\theta \in \Theta} |\phi_{\omega_{Q_\theta}}|^{-4}, \tag{B.1}$$

where $\mathcal{T}$ denotes a general subset of $(-\pi, \pi)^d$. We let Θ, the class of values of the parameter θ over which we search, be a compact set of vectors in a finite-dimensional Euclidean space. Below, the notation $C_2, C_3, \ldots$ and $c_1, c_2, \ldots$ will denote positive constants, large in the first instance and small in the second.

We assume that the white-noise variables ξ_j in (2.3) are such that:

> (a) $\xi_j(r)/\sigma_n$ are independent and identically distributed with zero mean and unit variance, for $r \in \mathcal{D}$, $j = 1, 2$ and all n; (b) $E|\xi_j(r)/\sigma_n|^{C_1} < \infty$, where $C_1 > 0$ is chosen sufficiently large, depending on c_1, c_2 and c_4. (B.2)

(The constants c_2 and c_4 are introduced in (B.5)(b) and (B.6)(b), respectively.) Property (B.2) requires that the noise variables be rescaled quantities with zero mean, unit variance and sufficiently many finite moments.

We also assume that the training sample sizes n_1 and n_2 satisfy:

> n_1 and n_2 are functions of n, and for some $c_1 > 0$, $n^{c_1} \le n_{\min} \equiv \min(n_1, n_2)$ for all n. (B.3)

This condition is equivalent to asking that training sample sizes increase at least polynomially fast as functions of the number of points in the grid $\mathcal{D}$.

Recall that $\#\mathcal{D} = (2n+1)^d$ denotes the number of vertices in the lattice $\mathcal{D}$, defined at (3.4). We model the distance between $T\mu_1$ and $T\mu_2$ by asking that:

> $\phi_{T(\mu_1 - \mu_2)} = \alpha_n \, \phi_{\omega_R} \, \phi_K$, where $|K(r)| \le C_2$ for all $r \in \mathcal{D}$ and $|\phi_K(t)| \le C_3 \, \#\mathcal{D}$ for all $t \in \mathcal{T}$. (B.4)

To understand this condition, it is helpful to interpret $K(r)$ as the value of a function, γ say, supported on $[-1, 1]$, at the point r/n, where n diverges to infinity in our asymptotic model. Let $\phi_\gamma(t)$ denote the continuous Fourier transform of γ, where $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$. For n large enough, we can expect $|\phi_K|$ to be well approximated by $(\#\mathcal{D})\,|\phi_\gamma|$. The last part of (B.4) reflects this property.

We assume too that the linear transformations Q_θ and R have the following symmetry and Hölder continuity properties:

> (a) the Fourier transforms $\phi_{\omega_{Q_\theta}}$ and ϕ_{ω_R} are real-valued; (b) the functional $|\phi_{\omega_{Q_\theta}}|$ satisfies $\sup_{t\in\mathcal{T}} \phi_{\omega_R}(t)^2\,|\phi_{\omega_{Q_\theta}}(t)^{-2} - \phi_{\omega_{Q_{\theta'}}}(t)^{-2}| \le C_4\,n^{C_5}\,\|\theta - \theta'\|^{c_2}$, uniformly in $\theta, \theta' \in \Theta$; (c) for all sufficiently large n, $\int_{\mathcal{T}} |\phi_K|^2\,|\phi_{\omega_R}|^4\,|\phi_{\omega_{Q_\theta}}|^{-4} \ge C_6\,(\#\mathcal{D})^2\,q$; (d) for all sufficiently large n, $\int_{\mathcal{T}} |\phi_K|^2\,|\phi_{\omega_R}|^2\,|\phi_{\omega_{Q_\theta}}|^{-2} \ge C_7\,\sigma_n^2\,(\#\mathcal{D})\,q$. (B.5)

Property (B.5)(a) asks that the functions ω_{Q_θ} and ω_R be symmetric. A longer proof can deal with cases of asymmetry, but symmetric models would almost always be used in practice. Condition (B.5)(b) asks that Q_θ and R satisfy a Hölder continuity condition in θ; in view of the factor n^{C_5}, and since c_2 can be arbitrarily small, it is quite mild. Properties (B.5)(c) and (B.5)(d) reflect the fact that $|\phi_K| \approx (\#\mathcal{D})\,|\phi_\gamma|$, as noted below (B.4).

We expect the transformation Q_θ to be reasonably close to R, and the following assumption responds to that property:

> (a) $\sup_{r\in\mathcal{D}} \sup_{\theta\in\Theta} |\omega_{R^2\,Q_\theta^{-2}}(r)| \le C_7$; (b) the quantity q, at (B.1), satisfies $n^{-c_3} \le q \le n^{c_4}$ for all sufficiently large n, where $0 < c_3 < 2d$, $0 < c_4 < c_1$ and c_1 is as in (B.3). (B.6)

Here $\omega_{R^2\,Q_\theta^{-2}}$ is the spread function associated with $R^2\,Q_\theta^{-2}$, that is

$$R^2\,Q_\theta^{-2}\zeta(r) = \sum_{s\in\mathbb{Z}} \omega_{R^2\,Q_\theta^{-2}}(s)\zeta(r+s).$$

We conclude with an example illustrating (B.5) and (B.6). Let p_0 be the probability mass function at (2.5) and let Q_θ denote the corresponding transformation, with associated Fourier transform $\phi_{\omega_{Q_\theta}}$ at (2.9). Then there exists $\rho_0 \in (0,1)$ and $D_0 \in (0,1)$ such that $D_0^d \le \phi_{\omega_{Q_\theta}}(t;\rho) \le D_0^{-d}$ for all $t \in (-\pi,\pi)$ and all $|\rho| \le \rho_0$. Put $\Theta = [-\rho_0, \rho_0]$, let $\theta_0 \in \Theta$, and let R be the transformation Q_{θ_0}. Then $|\phi_{\omega_R}(t)|\,|\phi_{\omega_{Q_\theta}}(t)|^{-1}$ is bounded away from zero and infinity uniformly in $t \in (-\pi,\pi)^d$ and in $\rho \in \Theta$, so (B.5)(c) and (B.5)(d) reduce to the assumption that $\int_{\mathcal{T}} |\phi_K|^2 \ge C\,(\#\mathcal{D})^2$, for some constant $C > 0$. That this condition holds for general choices of K follows from the fact that, in such cases, $\phi_K \approx (\#\mathcal{D})\,\phi_\gamma$ for a conventional, continuous Fourier transform ϕ_γ; see the argument below (3.4). Moreover, by taking the inverse Fourier transform we deduce that (B.6)(a) holds. Note too that these results are valid when $\mathcal{T} = (-\pi,\pi)^d$ and no smoothing is used.

B.2 Proof of Theorem 1

B.2.1 Step 1: Derivation of simpler criterion for classifier

The simpler criterion is at (B.8) below. To derive it, put $\bar{Y}_j = n_j^{-1} \sum_k Y_{kj}$ and let $\bar{Z}_{j;\theta}(r)$ be as defined immediately below (3.1). The Fourier transforms of Y and $\bar{Y}_j$ are, by (2.6),

$$\phi_Y(t) = \sum_{r \in \mathbb{Z}^d} Y(r) \, \exp\left(i \, r^{\mathrm{T}} t\right), \quad \phi_{\bar{Y}_j}(t) = n_j^{-1} \sum_{k=1}^{n_j} \sum_{r \in \mathbb{Z}^d} Y_{kj}(r) \, \exp\left(i \, r^{\mathrm{T}} t\right).$$

Likewise, the Fourier transform of $Z_\theta(r) - \bar{Z}_{j;\theta}(r)$ equals

$$\left\{\phi_Y(t) - \phi_{\bar{Y}_j}(t)\right\} \phi_{\omega_{Q_\theta}}(t)^{-1} \, I(t \in \mathcal{T}).$$

Therefore, since by (3.6) Z_θ is supported on $\mathcal{D}$, then by Parseval's identity,

$$\sum_{r \in \mathcal{D}} \left|Z_\theta(r) - \bar{Z}_{j;\theta}(r)\right|^2 = (2\pi)^{-d} \int_{\mathcal{T}} |\phi_Y - \phi_{\bar{Y}_j}|^2 \, |\phi_{\omega_{Q_\theta}}|^{-2},$$

and so the classifier represented by (3.1) is given equivalently by the algorithm: Assign Y to Π_1 if and only if

$$D(\theta) \equiv \int_{\mathcal{T}} |\phi_Y - \phi_{\bar{Y}_2}|^2 \, |\phi_{\omega_{Q_\theta}}|^{-2} - \int_{\mathcal{T}} |\phi_Y - \phi_{\bar{Y}_1}|^2 \, |\phi_{\omega_{Q_\theta}}|^{-2} > 0. \tag{B.7}$$

Write E for the expectation operator, put $\nu_j = T\mu_j$, write ν for the expected value of Y when the latter is drawn from one or other of Π_1 and Π_2, and put $\phi_\Delta = (1 - E)\phi_Y$ and $\phi_{\Delta_j} = (1 - E)\phi_{\bar{Y}_j}$, for $j = 1, 2$. Then,

$$\begin{aligned} D(\theta) &= \int_{\mathcal{T}} \Big[|\phi_{\nu-\nu_2}|^2 - |\phi_{\nu-\nu_1}|^2 + |\phi_{\Delta-\Delta_2}|^2 - |\phi_{\Delta-\Delta_1}|^2 \\ &\qquad + 2 \Big\{ \Re\phi_{\nu-\nu_2} \, \Re\phi_{\Delta-\Delta_2} + \Im\phi_{\nu-\nu_2} \, \Im\phi_{\Delta-\Delta_2} \\ &\qquad - \Re\phi_{\nu-\nu_1} \, \Re\phi_{\Delta-\Delta_1} - \Im\phi_{\nu-\nu_1} \, \Im\phi_{\Delta-\Delta_1} \Big\} \Big] \, |\phi_{\omega_{Q_\theta}}|^{-2} \\ &= D_1(\theta) + D_2(\theta) + 2\left\{D_{32}(\theta) - D_{31}(\theta)\right\} - 2\left\{D_{42}(\theta) - D_{41}(\theta)\right\}, \end{aligned} \tag{B.8}$$

where

$$\begin{aligned} D_1(\theta) &= \int_{\mathcal{T}} \left\{ |\phi_{\nu-\nu_2}|^2 - |\phi_{\nu-\nu_1}|^2 \right\} |\phi_{\omega_{Q_\theta}}|^{-2}, \\ D_2(\theta) &= \int_{\mathcal{T}} \left\{ |\phi_{\Delta-\Delta_2}|^2 - |\phi_{\Delta-\Delta_1}|^2 \right\} |\phi_{\omega_{Q_\theta}}|^{-2}, \end{aligned}$$

$$
\begin{aligned}
D_{3j}(\theta) &= \int_{\mathcal{T}} \left\{ \Re\phi_{\nu-\nu_j}\, \Re\phi_{\Delta} + \Im\phi_{\nu-\nu_j}\, \Im\phi_{\Delta} \right\} |\phi_{\omega_{Q_\theta}}|^{-2}\,, \\
D_{4j}(\theta) &= \int_{\mathcal{T}} \left\{ \Re\phi_{\nu-\nu_j}\, \Re\phi_{\Delta_j} + \Im\phi_{\nu-\nu_j}\, \Im\phi_{\Delta_j} \right\} |\phi_{\omega_{Q_\theta}}|^{-2}\,.
\end{aligned}
$$

B.2.2 Step 2. Formula for $D_1(\theta)$ and bound for $D_2(\theta)$

To simplify $D_1(\theta)$ take, for instance, $\nu = \nu_1$ and observe that, by (B.4),

$$
\int_{\mathcal{T}} |\phi_{\nu_1-\nu_2}|^2\, |\phi_{\omega_{Q_\theta}}|^{-2} = \alpha_n^2 \int_{\mathcal{T}} |\phi_K|^2\, |\phi_{\omega_R}|^2\, |\phi_{\omega_{Q_\theta}}|^{-2}\,. \tag{B.9}
$$

To bound $D_2(\theta)$, in a slight modification of the notation at (2.3) we write $\bar{Y}_j = \nu_j + R(\bar{\xi}_j)$ and $Y = \nu + R(\xi)$, where $\bar{\xi}_j = n_j^{-1} \sum_i \xi_{ji}$, and the random vectors ξ_{ji} (for $1 \le i \le n_j$ and $j = 1, 2$) and ξ are independent and identically distributed white noise; see (B.2). In this notation, $\Delta = R(\xi)$, $\Delta_j = R(\bar{\xi}_j)$,

$$
\phi_{\Delta} = \phi_{\xi}\, \phi_{\omega_R}\,, \quad \phi_{\Delta_j} = \phi_{\bar{\xi}_j}\, \phi_{\omega_R}\,, \tag{B.10}
$$

where, using (3.6),

$$
\phi_{\xi}(t) = \sum_{r \in \mathcal{D}} \xi(r)\, \exp\left(i r^{\mathrm{T}} t\right), \quad \phi_{\bar{\xi}_j}(t) = \sum_{r \in \mathcal{D}} \bar{\xi}_j(r)\, \exp\left(i\, r^{\mathrm{T}} t\right). \tag{B.11}
$$

Also,

$$
\begin{aligned}
\beta(t) &\equiv |\phi_{\Delta-\Delta_2}|^2 - |\phi_{\Delta-\Delta_1}|^2 \\
&= \Re\phi_{\Delta_1-\Delta_2}\, \Re\phi_{2\,\Delta-\Delta_1-\Delta_2} + \Im\phi_{\Delta_1-\Delta_2}\, \Im\phi_{2\,\Delta-\Delta_1-\Delta_2}\,.
\end{aligned} \tag{B.12}
$$

Defining

$$
q_0(t) = |\phi_{\omega_R}(t)|^2 \sup_{\theta \in \Theta} |\phi_{\omega_{Q_\theta}}(t)|^{-2}\,, \tag{B.13}
$$

we deduce from (B.10)–(B.12) that, uniformly in $\theta \in \Theta$,

$$
\begin{aligned}
\tfrac{1}{2}\, |D_2(\theta)| &\le \tfrac{1}{2} \int_{\mathcal{T}} |\beta|\, |\phi_{\omega_{Q_\theta}}|^{-2} \\
&\le \int_{\mathcal{T}} |\phi_{\Delta_1-\Delta_2}|\, |\phi_{2\,\Delta-\Delta_1-\Delta_2}|\, |\phi_{\omega_{Q_\theta}}|^{-2} \\
&\le \int_{\mathcal{T}} |\phi_{\bar{\xi}_1-\bar{\xi}_2}|\, |\phi_{2\,\xi-\bar{\xi}_1-\bar{\xi}_2}|\, q_0 \\
&\le \left\{ \int_{\mathcal{T}} |\phi_{\bar{\xi}_1-\bar{\xi}_2}|^2\, q_0 \right\}^{1/2} \left\{ \int_{\mathcal{T}} |\phi_{2\,\xi-\bar{\xi}_1-\bar{\xi}_2}|^2\, q_0 \right\}^{1/2},
\end{aligned} \tag{B.14}
$$

and moreover,

$$\sup_{t\in\mathcal{T}} E|\phi_{\bar{\xi}_j}(t)|^2 \le n_j^{-1}\,\sigma_n^2\,\#\mathcal{D}\,, \quad \sup_{t\in\mathcal{T}} E|\phi_{\xi}(t)|^2 \le \sigma_n^2\,\#\mathcal{D}\,. \tag{B.15}$$

Therefore, noting that q at (B.1), and q_0 at (B.13), satisfy $\int_{\mathcal{T}} q_0 \le \{(2\pi)^d\,q\}^{1/2}$ where the latter bound follows from the Cauchy-Schwarz inequality, we have:

$$\sup_{\theta\in\Theta} |D_2(\theta)| = O_p\big(\sigma_n^2\,n_{\min}^{-1/2}\,q\,\#\mathcal{D}\big)\,. \tag{B.16}$$

B.2.3 Step 3: Formula for $D_{3j}(\theta)$

As in step 2 we take $\nu = \nu_1$, in which case $D_{31} \equiv 0$ and

$$D_{32}(\theta) = \int_{\mathcal{T}} \Big\{\Re\phi_{\nu_1-\nu_2}\,\Re\phi_{\Delta} + \Im\phi_{\nu_1-\nu_2}\,\Im\phi_{\Delta}\Big\}\,|\phi_{\omega_{Q_\theta}}|^{-2}\,. \tag{B.17}$$

Using the fact that ϕ_{ω_R} is real-valued and letting a^* denote the complex conjugate of a complex number a, it can be deduced from (B.10), and the property $\phi_{\nu_1-\nu_2} = \alpha_n\,\phi_{\omega_R}\,\phi_K$ in (B.4), that

$$\begin{aligned}
&\alpha_n^{-1}\Big[\Re\big\{\phi_{\nu_1-\nu_2}(t)\big\}\,\Re\big\{\phi_{\Delta}(t)\big\} + \Im\big\{\phi_{\nu_1-\nu_2}(t)\big\}\,\Im\big\{\phi_{\Delta}(t)\big\}\Big] \\
&\quad = \Re\big\{\phi_K(t)\,\phi_{\omega_R}(t)^2\,\phi_{\xi}{}^*(t)\big\} \\
&\quad = \phi_{\omega_R}(t)^2\left[\left\{\sum_{r\in\mathcal{D}} K(r)\,\cos\big(r^{\mathrm{T}}t\big)\right\}\left\{\sum_{r\in\mathcal{D}} \xi(r)\,\cos\big(r^{\mathrm{T}}t\big)\right\}\right] \\
&\qquad\quad + \left\{\sum_{r\in\mathcal{D}} K(r)\,\sin\big(r^{\mathrm{T}}t\big)\right\}\left\{\sum_{r\in\mathcal{D}} \xi(r)\,\sin\big(r^{\mathrm{T}}t\big)\right\} \\
&\quad = \phi_{\omega_R}(t)^2 \sum_{r_1\in\mathcal{D}}\sum_{r_2\in\mathcal{D}} K(r_1)\,\xi(r_2)\cos\big\{(r_1-r_2)^{\mathrm{T}}t\big\}\,.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&(2\pi)^{-d}\,\alpha_n^{-1}\int_{\mathcal{T}} \Big\{\Re\phi_{\nu_1-\nu_2}\,\Re\phi_{\Delta} + \Im\phi_{\nu_1-\nu_2}\,\Im\phi_{\Delta}\Big\}\,|\phi_{\omega_{Q_\theta}}|^{-2} \\
&\qquad = (2\pi)^{-d}\sum_{r_1\in\mathcal{D}}\sum_{r_2\in\mathcal{D}} K(r_1)\,\xi(r_2)\int_{\mathcal{T}} \cos\big\{(r_1-r_2)^{\mathrm{T}}t\big\}\,\phi_{\omega_R}(t)^2\,|\phi_{\omega_{Q_\theta}}(t)|^{-2}\,dt \\
&\qquad = \sum_{r_1\in\mathcal{D}}\sum_{r_2\in\mathcal{D}} K(r_1)\,\xi(r_2)\,\omega_{R^2\,Q_\theta^{-2}}\,(r_1-r_2) \equiv Z_1(\theta)\,,
\end{aligned} \tag{B.18}$$

say. Note too that, defining A to be $\omega_{R^2\,Q_\theta^{-2}}$, the spread function associated with the transform $R^2\,Q_\theta^{-2}$, we have:

$$\begin{aligned}\tau(\theta)^2 &\equiv \mathrm{var}\{Z_1(\theta)\} = \sigma_n^2 \sum_{r_2} \Big\{ \sum_{r_1} K(r_1)\, A(r_1 - r_2) \Big\}^2 \\ &= (2\pi)^{-d}\, \sigma_n^2 \int_{\mathcal{T}} |\phi_K|^2\, |\phi_{\omega_R}|^4\, |\phi_{\omega_{Q_\theta}}|^{-4}\,. \end{aligned} \tag{B.19}$$

Defining $Z(\theta) = Z_1(\theta)/\tau(\theta)$, we deduce from (B.17) and (B.18) that, for all $\theta \in \Theta$,

$$D_{32}(\theta) = (2\pi)^d\, \alpha_n\, Z(\theta)\, \tau(\theta)\,. \tag{B.20}$$

B.2.4 Step 4: Bound for $D_{4j}(\theta)$

Again we take $\nu = \nu_1$, implying that $D_{41} \equiv 0$ and

$$D_{42}(\theta) = (2\pi)^d\, \alpha_n\, A_2(\theta) \tag{B.21}$$

where

$$\begin{aligned} A_j(\theta) &= (2\pi)^{-d}\, \alpha_n^{-1} \int_{\mathcal{T}} \Big\{ \Re\phi_{\nu_1-\nu_2}\, \Re\phi_{\Delta_j} + \Im\phi_{\nu_1-\nu_2}\, \Im\phi_{\Delta_j} \Big\}\, |\phi_{\omega_{Q_\theta}}|^{-2} \\ &= (2\pi)^{-d} \int_{\mathcal{T}} \Re\big(\phi_K\, \phi^*_{\bar\xi_j}\big)\, |\phi_{\omega_R}|^2\, |\phi_{\omega_{Q_\theta}}|^{-2}\,. \end{aligned} \tag{B.22}$$

The arguments leading to (B.18) and (B.19) imply that

$$A_j(\theta) = \sum_{r_1 \in \mathcal{D}} \sum_{r_2 \in \mathcal{D}} K(r_1)\, \bar\xi_j(r_2)\, \omega_{R^2\,Q_\theta^{-2}}\,(r_1 - r_2)\,, \tag{B.23}$$

$$\mathrm{var}\{A_j(\theta)\} = (2\pi)^{-d}\, \sigma_n^2\, n_j^{-1} \int_{\mathcal{T}} |\phi_K|^2\, |\phi_{\omega_R}|^4\, |\phi_{\omega_{Q_\theta}}|^{-4}\,. \tag{B.24}$$

Result (B.24) and assumption (B.4) imply that, with q defined as at (B.1),

$$\sup_{\theta\in\Theta} \mathrm{var}\{A_j(\theta)\} \le C_3^2\, \sigma_n^2\, n_j^{-1}\, q\, (\#\mathcal{D})^2\,. \tag{B.25}$$

Rosenthal's inequality implies that if $k \ge 1$ is an integer, and if $B_1(k) = E|\xi(r)/\sigma_n|^{2k}$ (a constant; see (B.2)(b)), then, for a constant $B_2(k)$ depending only on k, and with $B_3(k) = B_2(k)\,\{1 + B_1(k)\}$,

$$E\big\{\bar\xi_j(r_2)^{2k}\big\} \le B_2(k)\, \sigma_n^{2k}\, \big\{ n_j^{-k} + n_j^{1-2k}\, B_1(k) \big\} \le B_3(k)\, \sigma_n^{2k}\, n_j^{-k}\,. \tag{B.26}$$

The bound $\sup_{r\in\mathcal{D}} |K(r)| \le C_2$ in (B.4), and $\sup_{r\in\mathcal{D}} \sup_{\theta\in\Theta} |\omega_{R^2\,Q_\theta^{-2}}(r)| \le C_7$ in (B.6)(a), imply that

$$\sum_{r_1\in\mathcal{D}} \sum_{r_2\in\mathcal{D}} \bigl|K(r_1)\,\omega_{R^2\,Q_\theta^{-2}}(r_1-r_2)\bigr|^{2k} E\bigl\{\bar\xi_j(r_2)^{2k}\bigr\} \le (\#\mathcal{D})^2\,(C_2\,C_7)^{2k}\,B_3(k)\,\sigma_n^{2k}\,n_j^{-k}\,. \quad \text{(B.27)}$$

Property (B.23) expresses $A_j(\theta)$ as a sum of independent random variables, and hence, for any integer $k \in (1, \frac{1}{2}\,C_1]$ where C_1 is as in (B.2)(b), we can use Rosenthal's inequality again, together with (B.23) and (B.25)–(B.27), to show that

$$\begin{aligned} E|A_j(\theta)|^{2k} &\le B_2(k)\,\Bigl[\{\mathrm{var}A_j(\theta)\}^k \\ &\qquad + \sum_{r_1\in\mathcal{D}} \sum_{r_2\in\mathcal{D}} \bigl|K(r_1)\,\omega_{R^2\,Q_\theta^{-2}}(r_1-r_2)\bigr|^{2k} E\bigl\{\bar\xi_j(r_2)^{2k}\bigr\}\Bigr] \\ &\le B_4(k)\,\Bigl[\bigl\{\sigma_n^2\,n_j^{-1}\,q\,(\#\mathcal{D})^2\bigr\}^k + (\#\mathcal{D})^2\,\bigl(\sigma_n^2\,n_j^{-1}\bigr)^k\Bigr] \\ &\le B_5(k)\,\bigl\{\sigma_n^2\,n_j^{-1}\,q\,(\#\mathcal{D})^2\bigr\}^k\,, \end{aligned} \quad \text{(B.28)}$$

where q is as at (B.1) and the last inequality holds if k (and hence C_1 in (B.2)(b)) is taken sufficiently large. To obtain the last inequality in (B.28), we used the lower bound $n^{-c_3} \le q$ (where $c_3 < 2d$) in (B.6)(b) to ensure that $(\#\mathcal{D})^2\,(\sigma_n^2\,n_j^{-1})^k = O[\{\sigma_n^2\,n_j^{-1}\,q\,(\#\mathcal{D})^2\}^k]$ for sufficiently large k. The argument here requires $c_3 \le 2d\,(1-k^{-1})$.

By (B.3) and (B.6)(b), $n_j^{-1}\,(q+1) \le n_{\min}^{-1}\,(q+1) = O(n^{-(c_1-c_4)})$, where c_4 is as in (B.6)(b) and $c \equiv c_1 - c_4 > 0$. Hence, by (B.28) and Markov's inequality,

$$\sup_{\theta\in\Theta} P\bigl\{|A_j(\theta)| > \epsilon\,\sigma_n\,n_{\min}^{-1/2}\,q^{1/2}\,\#\mathcal{D}\bigr\} \le B_6(k)\,\bigl(\epsilon\,n^c\bigr)^{-k}$$

for all $\epsilon > 0$. Therefore, if $\Theta_n \subseteq \Theta$ satisfies

$$\#\Theta_n = o(n^{kc}) \quad \text{(B.29)}$$

then, for all $\epsilon > 0$,

$$P\biggl\{\sup_{\theta\in\Theta_n} |A_j(\theta)| > \epsilon\,\sigma_n\,n_{\min}^{-1/2}\,q^{1/2}\,\#\mathcal{D}\biggr\} \to 0\,. \quad \text{(B.30)}$$

Recall that Θ is a compact subset of a finite-dimensional Euclidean space. Given $B_6 > 0$ we can choose $B_7 > 0$, depending on B_6 and on C_5 and c_2 in (B.5)(b), so large that for all sufficiently large n there exists a set $\Theta_n \subseteq \Theta$ such that (B.29) holds if k is chosen sufficiently

large (i.e. if C_1 in (B.2)(b) is taken sufficiently large), and the following property obtains: if $\theta' \in \Theta_n$ denotes the value nearest to $\theta \in \Theta$ then

$$\sup_{t \in \mathcal{T}} \phi_{\omega_R}(t)^2 \, |\phi_{\omega_{Q_\theta}}(t)^{-2} - \phi_{\omega_{Q_{\theta'}}}(t)^{-2}| \le n^{-B_7} \, .$$

This bound, together with (B.22), implies that, with C_3 as in (B.4),

$$\begin{aligned} \sup_{\theta \in \Theta} |A_j(\theta) - A_j(\theta')| &\le (2\pi)^{-d} \, n^{-B_7} \int_{\mathcal{T}} \bigl| \Re\bigl(\phi_K \, \phi^*_{\bar{\xi}_j}\bigr) \bigr| \\ &\le (2\pi)^{-d} \, C_3 \, (\#\mathcal{D}) \, n^{-B_7} \int_{\mathcal{T}} |\phi_{\bar{\xi}_j}| \\ &= O_p\bigl\{ \sigma_n \, n_j^{-1/2} \, n^{-B_7} \, (\#\mathcal{D})^{3/2} \bigr\} \, , \end{aligned} \tag{B.31}$$

where the last identity follows from the bound $\int_{\mathcal{T}} |\phi_{\bar{\xi}_j}| = O_p\{\sigma_n \, n_j^{-1/2} \, (\#\mathcal{D})^{1/2}\}$. Since $\#\mathcal{D} = O(n^d)$ (see the definition (3.4) of $\mathcal{D}$, and the first paragraph of this section) then, by taking B_7 sufficiently large, we can deduce from (B.30) and (B.31) that

$$\sup_{\theta \in \Theta} |A_j(\theta)| = O_p\bigl(\sigma_n \, n_{\min}^{-1/2} \, q^{1/2} \, \#\mathcal{D}\bigr) \, . \tag{B.32}$$

Together, (B.21) and (B.32) imply that

$$\sup_{\theta \in \Theta} |D_{42}(\theta)| = O_p\bigl(\alpha_n \, \sigma_n \, n_{\min}^{-1/2} \, q^{1/2} \, \#\mathcal{D}\bigr) \, . \tag{B.33}$$

B.2.5 Step 5: Formula for classification error

Combining (B.8), (B.9), (B.16), (B.19), (B.20) and (B.33), and bearing in mind that these properties were derived under the assumption that the new variable Y came from population Π_1 (that is, $\nu = \nu_1$), we deduce that when Y comes from Π_1,

$$\begin{aligned} D(\theta) =& \alpha_n^2 \int_{\mathcal{T}} |\phi_K|^2 \, |\phi_{\omega_R}|^2 \, |\phi_{\omega_{Q_\theta}}|^{-2} + \alpha_n \, \sigma_n \left\{ (2\pi)^d \int_{\mathcal{T}} |\phi_K|^2 \, |\phi_{\omega_R}|^4 \, |\phi_{\omega_{Q_\theta}}|^{-4} \right\}^{1/2} Z(\theta) \\ &+ O_p\Bigl(\sigma_n^2 \, n_{\min}^{-1/2} \, q \, \#\mathcal{D} + \alpha_n \, \sigma_n \, n_{\min}^{-1/2} \, q^{1/2} \, \#\mathcal{D} \Bigr) \, , \end{aligned} \tag{B.34}$$

uniformly in $\theta \in \Theta$.

The random variable $Z(\theta)$ in (B.34) has, by construction (see (B.19) and the line above (B.20)), zero mean and unit variance. We shall use Lindeberg's central limit theorem to prove that $Z(\theta)$ is asymptotically normal $\mathrm{N}(0,1)$. In view of (B.2), (B.18) and (B.19), to establish

that Lindeberg's condition (see e.g. Chung 1974, p. 205) holds uniformly in $\theta \in \Theta$ it suffices to prove that for some $k > 0$, and with $P = \omega_{R^2} Q_\theta^{-2}$,

$$\sup_{\theta \in \Theta} \frac{\sum_{r_2 \in \mathcal{D}} |\sum_{r_1 \in \mathcal{D}} K(r_1)\, P(r_1 - r_2)|^{k+2}}{\left\{ \int_{\mathcal{T}} |\phi_K|^2\, |\phi_{\omega_R}|^4\, |\phi_{\omega_{Q_\theta}}|^{-4} \right\}^{(k+2)/2}} \to 0 \tag{B.35}$$

as $n \to \infty$. Now,

$$\begin{aligned} \left| \sum_{r_1 \in \mathcal{D}} K(r_1)\, P(r_1 - r_2) \right|^2 &\le \left\{ \sum_{r \in \mathcal{D}} K(r)^2 \right\} \left\{ \sum_{r \in \mathbb{Z}^d} P(r)^2 \right\} \\ &\le C_2^2\, (\#\mathcal{D})\, (2\pi)^d \int_{\mathcal{T}} |\phi_{\omega_R}|^4\, |\phi_{\omega_{Q_\theta}}|^{-4} \\ &= C_2^2\, (2\pi)^d\, (\#\mathcal{D})\, q = B_8\, (\#\mathcal{D})^{-1} \cdot (\#\mathcal{D})^2\, q \\ &\le B_9\, (\#\mathcal{D})^{-1} \int_{\mathcal{T}} |\phi_K|^2\, |\phi_{\omega_R}|^4\, |\phi_{\omega_{Q_\theta}}|^{-4}\,, \end{aligned} \tag{B.36}$$

where C_2 is as in (B.4), B_8 and B_9 do not depend on θ, and the final inequality in (B.36) follows from (B.5)(c). Hence, the left-hand side of (B.35) is bounded above by $B_9\, (\#\mathcal{D})^{1-(k+2)/2}$, which converges to zero if $k > 0$. It can be proved from this property that the following uniform version of the central limit theorem holds:

$$\sup_{\theta \in \Theta} \sup_{-\infty < x < \infty} \left| P\{Z(\theta) \le x\} - \Phi(x) \right| \to 0\,. \tag{B.37}$$

Note that, in view of (B.5)(c), the multiplier of $Z(\theta)$ on the right-hand side of (B.34) is bounded below by a constant multiple of $\alpha_n\, \sigma_n\, q^{1/2}\, \#\mathcal{D}$, and so is of strictly larger order than the remainder term $O_p(\alpha_n\, \sigma_n\, n_{\min}^{-1/2}\, q^{1/2}\, \#\mathcal{D})$ in (B.34); and, by (B.5)(d), the first term on the right-hand side of (B.34) is bounded below by a constant multiple of $\sigma_n^2\, q\, \#\mathcal{D}$, and so is of strictly larger order than the remainder term $O_p(\sigma_n^2\, n_{\min}^{-1/2}\, q\, \#\mathcal{D})$ in (B.34). Combining the results in the previous sentence with (B.34) and (B.37) we deduce that, uniformly in $\theta \in \Theta$,

$$\begin{aligned} D(\theta) &= \{1 + o_p(1)\}\, \alpha_n^2 \int_{\mathcal{T}} |\phi_K|^2\, |\phi_{\omega_R}|^2\, |\phi_{\omega_{Q_\theta}}|^{-2} \\ &\quad + \{1 + o_p(1)\}\, \alpha_n\, \sigma_n \left\{ (2\pi)^d \int_{\mathcal{T}} |\phi_K|^2\, |\phi_{\omega_R}|^4\, |\phi_{\omega_{Q_\theta}}|^{-4} \right\}^{1/2} Z(\theta)\,. \end{aligned} \tag{B.38}$$

Together, (B.37) and (B.38) imply that, with u_n defined as at (3.7), we have:

$$\sup_{\theta \in \Theta} \left| P\{D(\theta) < 0 \,\big|\, Y \in \Pi_1\} - \Phi(-u_n) \right| \to 0$$

as $n \to \infty$, and similarly,

$$\sup_{\theta \in \Theta} \left| P\{D(\theta) > 0 \;\big|\; Y \in \Pi_2\} - \Phi(-u_n) \right| \to 0 \,.$$

This establishes (3.8).

B.3 Proof of Theorem 2

Formula (B.38), for $D(\theta)$ defined at (B.7), remains correct if we interpret it as applying in the case where Y is drawn from the training sample $\mathcal{Y}_1$, and if we regard $Z(\theta)$ as being computed from this particular Y. (Recall that $Z(\theta)$ is calculated from Y alone; it has no dependence on $\mathcal{Y}$.) Suppose we can extend (B.38) so that it applies not just to one such choice of Y, but uniformly over all n_1 such choices, giving values $(D_{[\ell]}(\theta), Z_\ell(\theta))$, for $1 \le \ell \le n_1$, of the pairs $(D(\theta), Z(\theta))$. Here n_1 is reduced to $n_1 - 1$, but n_2 remains unchanged, and we have

$$\begin{aligned} D_{[\ell]}(\theta) \;&=\; \{1 + o_p(1)\}\, \alpha_n^2 \int_{\mathcal{T}} |\phi_K|^2\, |\phi_{\omega_R}|^2\, |\phi_{\omega_{Q_\theta}}|^{-2} \\ &\quad + \{1 + o_p(1)\}\, \alpha_n\, \sigma_n \left\{ (2\pi)^d \int_{\mathcal{T}} |\phi_K|^2\, |\phi_{\omega_R}|^4\, |\phi_{\omega_{Q_\theta}}|^{-4} \right\}^{1/2} Z_\ell(\theta)\,, \end{aligned} \tag{B.39}$$

where the $o_p(1)$ remainders are of the stated orders uniformly in $1 \le \ell \le n_1$ as well as in $\theta \in \Theta$.

Computing the indicator function $I\{D_{[\ell]}(\theta) > 0\}$ on the left-hand side of (B.39), averaging over ℓ, and applying a law of large numbers for the sum over $\ell = 1, \ldots, n_1$ of the independent random variables $I\{Z_\ell(\theta) > u_n\}$, we obtain:

$$\frac{1}{n_1} \sum_{\ell=1}^{n_1} I\{D_{[\ell]}(\theta) > 0\} = \Phi(-u_n) + o_p(1)\,,$$

uniformly in $\theta \in \Theta$. Multiplying this by π_1, computing its counterpart when Y is drawn from $\mathcal{Y}_2$ rather than $\mathcal{Y}_1$, and adding them together, we obtain exactly, on the left-hand side, $\widehat{e}(\theta)$, and on the right-hand side, $1 - \Phi(u_n) + o_p(1)$, uniformly in $\theta \in \Theta$. This is equivalent to (3.9) and so completes the proof of Theorem 2.

It remains to verify that (B.39) holds uniformly in $1 \le \ell \le n_1$ and $\theta \in \Theta$. In describing how this is done we refer to formula (B.8) for $D(\theta)$, taking a subscript ℓ (given above in square brackets, so as not to cause confusion with notation below (B.8)) to be understood.

The quantity $D_1(\theta)$ does not alter, and the bound (B.14) for $\frac{1}{2}\,|D_2(\theta)|$ remains valid except that the pair $(\xi, \bar{\xi}_1)$ now has n_1 different versions $(\xi_\ell, \bar{\xi}_{1\ell})$ say, each producing its own $D_2(\theta)$. (There is only one version of $\bar{\xi}_2$.) However, using Parseval's inequality to bound $E|\phi_{\xi_\ell}|^{2k}$ and $E|\phi_{\bar{\xi}_{1\ell}}|^{2k}$ for any $k \geq 1$, and noting that (B.2)(b) permits us to take k as large as needed, we obtain:

$$E\Big\{\max_{1\leq\ell\leq n_1}|\phi_{\bar{\xi}_{1\ell}}(t)|^2\Big\} \leq \Big[E\Big\{\sum_{\ell=1}^{n_1}|\phi_{\bar{\xi}_{1\ell}}(t)|^{2k}\Big\}\Big]^{1/k} = n_1^{1/k}\,\big\{E|\phi_{\bar{\xi}_{11}}(t)|^{2k}\big\}^{1/k} = O\big(n_1^{-(k-1)/k}\,\sigma_n^2\,\#\mathcal{D}\big)\,,$$

uniformly in $t \in \mathcal{T}$. This bound replaces the first part of (B.15), and the second part of (B.15) can similarly be replaced by $O(n_1^{1/k}\,\sigma_n^2\,\#\mathcal{D})$. This argument shows that (B.16) continues to hold, this time uniformly over all leave-one-out cases, provided we replace $n_{\min}^{-1/2}$ on the right-hand side by $n_{\min}^{\epsilon-(1/2)}$, where $\epsilon > 0$ can be made as small as we like by choosing C_1, in (B.2)(a), sufficiently large. That change to (B.16) requires us to replace $n_{\min}^{-1/2}$ by $n_{\min}^{\epsilon-(1/2)}$ in the $O_p(\sigma_n^2\,n_{\min}^{-1/2}\,q\,\#\mathcal{D})$ term on the right-hand side of (B.34), but this alteration does not influence any of the subsequent arguments in step 5.

The term $D_{3j}(\theta)$, which depends only on Y (or, in the present circumstance, on an observation drawn from $\mathcal{Y}_1$) can be dealt with exactly as before. Moreover, an upper bound to $|D_{4j}(\theta)|$ can be derived, uniformly over all leave-one-out choices, by following the argument given in the previous paragraph for $D_2(\theta)$. In this way it can be proved that the contribution made by $D_{4j}(\theta)$ to $D(\theta)$ can be accommodated in the $\alpha_n\,\sigma_n\,n_{\min}^{-1/2}\,q^{1/2}\,\#\mathcal{D}$ remainder term in (B.34), provided that $n_{\min}^{-1/2}$ on the right-hand side is replaced by $n_{\min}^{\epsilon-(1/2)}$, again for arbitrarily small $\epsilon > 0$ and uniformly over all leave-one-out cases; and that this has no bearing on later arguments in step 5.

Additional References

Chung, K. L. (1974), *A Course in Probability Theory* (2nd ed.), New York: Academic Press.