Supplementary Material to "Density Level Sets: Asymptotics, Inference, and Visualization"

Yen-Chi Chen*
Department of Statistics, University of Washington
and
Christopher R. Genovese†
Department of Statistics, Carnegie Mellon University
and
Larry Wasserman‡
Department of Statistics, Carnegie Mellon University

August 18, 2016

Abstract

This document contains proofs of theorems in 'Density Level Sets: Asymptotics, Inference, and Visualization'.

Keywords: Nonparametric inference, asymptotic theory, level set clustering, anomaly detection, visualization

*Supported by the William S. Dietrich II Presidential Ph.D. Fellowship and DOE grant number DE-FOA-0000918

†Supported by DOE grant number DE-FOA-0000918 and NSF grant number DMS-1208354

‡Supported by NSF grant number DMS-1208354

A Proofs

Theorem 8 (Theorem 2 in Cuevas et al. (2006))**.** *Assume (K1–2) and (G), then we have*

$$\mathsf{Haus}\left(\widehat{D}_h, D_h\right) = O(\|\widehat{p}_h - p_h\|_{0,\max}).$$

Theorem 9 (Talagrand's inequality; version of Theorem 12 in Chen et al. (2014a))**.** *Assume (K1–2), then for each $t > 0$ there exists some n_0 such that whenever $n > n_0$, we have*

$$\mathbb{P}\left(\|\widehat{p}_h - p_h\|^*_{\ell,\max} > t\right) \leq (\ell+1)e^{-tnh^{d+2\ell}A_1},$$

for some constant A_1 and $\ell = 0, 1, 2$. Moreover,

$$\mathbb{E}\left(\|\widehat{p}_h - p_h\|^*_{2,\max}\right) = O\left(\sqrt{\frac{\log n}{nh^{d+4}}}\right).$$

Proof for Lemma 1. We first prove the lower bound for $\mathsf{reach}(D_h)$ and then we will prove the additional assertions.

Part 1: Lower bound on reach. We prove this by contradiction. Take x near D_h such that

$$d(x, D_h) < \left(\frac{\delta_0}{2}, \frac{g_0}{\|p_h\|^*_{2,\max}}\right). \tag{1}$$

We assume that x has two projections onto D_h, denoted as b and c.

Since $b, c \in D_h$, $p_h(b) - \lambda = p_h(c) - \lambda = 0$ so that $p_h(b) - p_h(c) = 0$. Now by Taylor's theorem

$$\begin{aligned}\|(b-c)^T\nabla p_h(b)\| &= \|p_h(b) - p_h(c) - (b-c)^T\nabla p_h(b)\| \\ &\leq \frac{1}{2}\|b-c\|^2\|p_h\|_{2,\max}.\end{aligned} \tag{2}$$

By the definition of projection, we can find a constant $t_b \in \mathbb{R}$ such that $x - b = t_b\nabla p_h(b)$. Together with (2),

$$\begin{aligned}2|(b-c)^T(x-b)| &= 2|(b-c)^T\nabla p_h(b)t_b| \\ &\leq \|(b-c)^T\nabla p_h(b)\||t_b| \\ &\leq \|p_h\|_{2,\max}\|b-c\|^2|t_b|.\end{aligned} \tag{3}$$

Since both b and c are projection points from x onto D_h,

$$\|x - b\| = \|x - c\|.$$

Thus, we have

$$\begin{aligned} 0 &= \|x - c\|^2 - \|x - b\|^2 \\ &= \|b - c\|^2 + 2(b - c)^T(x - b) \\ &\geq \|b - c\|^2 - \|p_h\|_{2,\max}\|b - c\|^2|t_b| \\ &= \|b - c\|^2(1 - \|p_h\|_{2,\max}|t_b|). \end{aligned} \tag{4}$$

Recall that $d(x, D_h) \leq \frac{g_0}{\|p_h\|_{2,\max}}$ and by Taylor's theorem,

$$\frac{g_0}{\|p_h\|_{2,\max}} > d(x, D_h) = \|x - b\| = \|t_b \nabla p_h(b)\| = |t_b|\|\nabla p_h(b)\| \geq |t_b| g_0 \tag{5}$$

so that $|t_b|\|p_h\|_{2,\max} < 1$. Note that the lower bound g_0 in the last inequality is because $d(x, D_h) < \frac{\delta_0}{2}$ so it follows from assumption (G). Plugging in this result into the last equality of (4), we conclude that $\|b - c\| = 0$. This shows $b = c$ so that we have a unique projection. Thus, whenever $d(x, D_h) < \left(\frac{\delta_0}{2}, \frac{g_0}{\|p_h\|^*_{2,\max}}\right)$, we have a unique projection onto D_h and thus we have proved the lower bound on reach.

Part 2: The three assertions. The first assertion is trivially true when $\|p_h - q\|^*_{2,\max}$ is sufficiently small since assumption (G) only involves gradients (first derivatives).

The second assertion follows from the lower bound on reach. By assertion 1, (G) holds for q. And the lower bound on reach is bounded by gradient and second derivatives so that we have the prescribed bound.

The third assertion follows from Theorem 1 in Chazal et al. (2007) which states that if two $d - 1$ dimensional smooth manifolds M_1 and M_2 have Hausdorff distance being less than $(2 - \sqrt{2}) \min\{\mathsf{reach}(M_1), \mathsf{reach}(M_2)\}$, then M_1 and M_2 are normal compatible to each other. Now by Theorem 8, the Hausdorff distance between D_h and $D(q)$ is at rate $O(\|p_h - q\|_{1,\max})$ so that this assertion is true when $\|p_h - q\|_{2,\max}$ is sufficiently small. $\square$

Proof of Lemma 2. Let $x \in D_h$. We define $\Pi(x) \in D_h$ to be the projected point onto $\widehat{D}_h$. By Lemma 1 and Theorem 8, when $\|\widehat{p}_h - p_h\|^*_{2,\max} \to 0$, $\mathsf{Haus}(D_h, \widehat{D}_h) \xrightarrow{P} 0$ so that $\Pi(x)$ is unique. Thus, we assume $\Pi(x)$ is unique.

Now since $\Pi(x) \in \widehat{D}_h$ and $x \in D_h$, $\widehat{p}_h(\Pi(x)) - p_h(x) = 0$. Thus, by Taylor's theorem

$$\begin{aligned}\widehat{p}_h(x) - p_h(x) &= \widehat{p}_h(x) - \widehat{p}_h(\Pi(x)) \\ &= (x - \Pi(x))^T(\nabla \widehat{p}_h(\Pi(x)) + O_{\mathbb{P}}(\|x - \Pi(x)\|)).\end{aligned} \tag{6}$$

Note that $x - \Pi(x)$ is normal to $\widehat{D}_h$ at $\Pi(x)$ so that it points toward the same direction as $\nabla \widehat{p}_h(\Pi(x))$. Thus, (6) can be rewritten as

$$\widehat{p}_h(x) - p_h(x) = \|x - \Pi(x)\| \big(\|\nabla \widehat{p}_h(\Pi(x))\| + O_{\mathbb{P}}(\|x - \Pi(x)\|)\big). \tag{7}$$

By Taylor's theorem, $\nabla \widehat{p}_h(\Pi(x))$ is close to $\nabla p_h(x)$ in the sense that

$$\nabla \widehat{p}_h(\Pi(x)) = \nabla p_h(x) + O(\|\widehat{p}_h - p_h\|^*_{1,\max}). \tag{8}$$

In addition, $O(\|x - \Pi(x)\|)$ is bounded by $O(\mathsf{Haus}(\widehat{D}_h, D_h))$ which is at rate $O(\|\widehat{p}_h - p_h\|^*_{1,\max})$ due to Theorem 8. Putting this together with (7), we conclude

$$\begin{aligned}\widehat{p}_h(x) - p_h(x) &= \|x - \Pi(x)\| \big(\|p_h(x)\| + O(\|\widehat{p}_h - p_h\|^*_{1,\max})\big) \\ &= d(x, \widehat{D}_h)\big(\|p_h(x)\| + O(\|\widehat{p}_h - p_h\|^*_{1,\max})\big).\end{aligned} \tag{9}$$

Note that the left hand side can be written as

$$\widehat{p}_h(x) - p_h(x) = \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) - \frac{1}{h^d}\mathbb{E}\left(K\left(\frac{x - X_i}{h}\right)\right) = \frac{1}{\sqrt{nh^d}}\mathbb{G}_n(\tilde{f}_x), \tag{10}$$

where $\tilde{f}_x(y) = K\left(\frac{x-y}{h}\right)$. After plugging (10) into the left hand side of (9), dividing both side by $\|p_h(x)\|$ and setting $f_x(y) = \frac{\tilde{f}_x(y)}{\sqrt{h^d}\|p_h(x)\|}$, we obtain

$$\frac{\frac{1}{\sqrt{nh^d}}\mathbb{G}_n(f_x) - d(x, \widehat{D}_h)}{d(x, \widehat{D}_h)} = O(\|\widehat{p}_h - p_h\|^*_{1,\max}). \tag{11}$$

This holds uniformly for all $x \in D_h$ and note that the definition of $\mathcal{F}$ is

$$\mathcal{F} = \left\{ f_x(y) \equiv \frac{1}{\sqrt{h^d}\|\nabla p_h(x)\|} K\left(\frac{x - y}{h}\right) : x \in D_h \right\}.$$

So we conclude

$$\sup_{x\in D_h}\left|\frac{\frac{1}{\sqrt{nh^d}}\mathbb{G}_n(f_x)-d(x,\widehat{D}_h)}{d(x,\widehat{D}_h)}\right| = O(\|\widehat{p}_h-p_h\|^*_{1,\max}).$$

$\square$

PROOF FOR THEOREM 3. The proof for Theorem 3 follows the same procedure as the proof of Theorem 6 in Chen et al. (2014b). The proof contains two parts: Gaussian approximation and anti-concentration.

Part 1: Gaussian approximation. Basically, we will show that

$$\sqrt{nh^d}\,\mathsf{Haus}(\widehat{D}_h, D_h) \approx \sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)| \approx \sup_{f\in\mathcal{F}}|\mathbb{B}(f)|,$$

where $\mathbb{B}$ is a Gaussian process defined in (13) of the original paper.

First, when $\|\widehat{p}_h - p_h\|$ is sufficiently small, $\widehat{D}_h$ and D_h are normal compatible to each other by Lemma 1. Then by the property of normal compatible,

$$\sup_{x\in D_h} d(x,\widehat{D}_h) = \mathsf{Haus}(\widehat{D}_h, D_h). \tag{12}$$

Thus, the difference

$$\begin{aligned}\left|\sqrt{nh^d}\,\mathsf{Haus}(\widehat{D}_h, D_h) - \sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|\right| &= \left|\sqrt{nh^d}\sup_{x\in D_h} d(x,\widehat{D}_h) - \sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|\right|\\ &\leq \frac{\sup_{x\in D_h}\left|\frac{1}{\sqrt{nh^d}}\mathbb{G}_n(f_x)-d(x,\widehat{D}_h)\right|}{\frac{1}{\sqrt{nh^d}}}\\ &= \sup_{x\in D_h}\left|\frac{\frac{1}{\sqrt{nh^d}}\mathbb{G}_n(f_x)-d(x,\widehat{D}_h)}{d(x,\widehat{D}_h)}\right| O_{\mathbb{P}}(1)\\ &= O(\|\widehat{p}_h-p_h\|^*_{1,\max}).\end{aligned} \tag{13}$$

Note that the last two inequality follows from the fact that $d(x,\widehat{D}_h)\leq O_{\mathbb{P}}(\frac{1}{\sqrt{nh^d}})$. By Theorem 9 the above result implies,

$$\mathbb{P}\left(\left|\sqrt{nh^d}\,\mathsf{Haus}(\widehat{D}_h, D_h) - \sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|\right| > t\right) \leq 2e^{-tnh^{d+2}A_2} \tag{14}$$

for some constant A_2.

Now by Corollary 2.2 in Chernozhukov et al. (2014c), there exists some random variable $\mathbf{B} \stackrel{d}{=} \sup_{f\in\mathcal{F}} |\mathbb{B}(f)|$ such that for all $\gamma \in (0,1)$ and n is sufficiently large,

$$\mathbb{P}\left(\left|\sup_{f\in\mathcal{F}} |\mathbb{G}_n(f)| - \mathbf{B}\right| > A_3 \frac{\log^{2/3}(n)}{\gamma^{1/3}(nh^d)^{1/6}}\right) \leq A_4\gamma. \tag{15}$$

Note that this result basically follows from the same derivation of Proposition 3.1 in Chernozhukov et al. (2014c) with the fact that $g \equiv 1$ in their definition.

Combining equations (14) and (15) and pick $t = 1/\sqrt{nh^{d+2}}$, we have that for n is sufficiently large and $\gamma \in (0,1)$,

$$\mathbb{P}\left(\left|\sqrt{nh^d}\mathsf{Haus}(\widehat{D}_h, D_h) - \mathbf{B}\right| > A_3 \frac{\log^{2/3}(n)}{\gamma^{1/3}(nh^d)^{1/6}} + \frac{1}{\sqrt{nh^{d+2}}}\right) \leq A_4\gamma + 2e^{-\sqrt{nh^{d+2}}A_2}. \tag{16}$$

Part 2: Anti-concentration. To obtain the desired Berry-Esseen bound, we apply the anti-concentration inequality in Chernozhukov et al. (2014c) and Chernozhukov et al. (2014a).

Lemma 10 (Modification of Lemma 2.3 in Chernozhukov et al. (2014c))**.** *Let* $\mathbf{B} \stackrel{d}{=} \sup_{f\in\mathcal{F}} |\mathbb{B}(f)|$, *where* $\mathbb{B}$ *and* $\mathcal{F}$ *are defined as the above. Assume (K1-2) and that there exists a random variable* Y *such that* $\mathbb{P}(|Y - \mathbf{B}| > \eta) < \delta(\eta)$. *Then*

$$\sup_t |\mathbb{P}(Y < t) - \mathbb{P}(\mathbf{B} < t)| \leq A_5\mathbb{E}(\mathbf{B})\eta + \delta(\eta)$$

for some constant A_5.

It is easy to verify that assumptions (K1-2) imply the assumptions (A1-3) in Chernozhukov et al. (2014c) so that the result follows. Note that in the original Lemma 2.3 in Chernozhukov et al. (2014c), $\mathbb{E}(\mathbf{B})$ should be replaced by $\mathbb{E}(\mathbf{B}) + \log\eta$. However, $\mathbb{E}(\mathbf{B}) = O(\sqrt{\log n})$ due to Dudley's inequality for Gaussian process (Van Der Vaart and Wellner, 1996) and later we will find that $\log\eta$ is also at this rate so we ignore $\log\eta$.

From Lemma 10 and equation (16), there exists some constant A_6 such that

$$\begin{aligned}
\sup_t \left| \mathbb{P}\left(\sqrt{nh^d}\mathsf{Haus}(\widehat{D}_h, D_h) < t\right) - \mathbb{P}\left(\sup_{f\in\mathcal{F}} |\mathbb{B}(f)| < t\right) \right| \\
\leq A_5 \mathbb{E}(\mathbf{B}) \left(A_3 \frac{\log^{2/3}(n)}{\gamma^{1/3}(nh^d)^{1/6}} + \frac{1}{\sqrt{nh^{d+2}}} \right) + A_4\gamma + 2e^{-\sqrt{nh^{d+2}}A_2} \quad (17)\\
\leq A_6 \left(A_3 \frac{\log^{7/6}(n)}{\gamma^{1/3}(nh^d)^{1/6}} + \sqrt{\frac{\log n}{nh^{d+2}}} \right) + A_4\gamma + 2e^{-\sqrt{nh^{d+2}}A_2}.
\end{aligned}$$

Now pick $\gamma = \left(\frac{\log^7 n}{nh^d}\right)^{1/8}$ and use the fact that $\frac{1}{\sqrt{nh^{d+2}}}$ and $2e^{-\sqrt{nh^{d+2}}A_2}$ converges faster than the other terms; we obtain the desired rate. $\square$

Proof for Theorem 4. This proof follows the same strategy for the proof of Theorem 7 in Chen et al. (2014b). We prove the Berry-Esseen type bound first and then show that the coverage is consistent. We prove the Berry-Esseen bound in two simple steps: Gaussian approximation and support approximation.

Let $\mathcal{X}_n = \{(X_1, \cdots, X_n) : \|\widehat{p}_h - p_h\|^*_{2,\max} \leq \eta_0\}$ for some small η_0 so that whenever our data is within $\mathcal{X}_n$, (G) holds for $\widehat{p}_h$. By Lemma 1, such an η_0 exists and by Theorem 9 we have $\mathbb{P}(\mathcal{X}_n) \geq 1 - 3^{-nh^{d+4}\tilde{A}_0}$ for some constant $\tilde{A}_0$. Thus, we assume our original data $X_1, \cdots, X_n$ is within $\mathcal{X}_n$.

Step 1: Gaussian approximation. Let $\widehat{\mathbb{P}}_n$ and $\widehat{\mathbb{P}}^*_n$ be the empirical measure and the bootstrap empirical measure. A crucial observation is that for a function $f_x(y) = K\left(\frac{x-y}{h}\right)$,

$$\widehat{\mathbb{P}}_n(f_x) = \int K\left(\frac{x-y}{h}\right) d\widehat{\mathbb{P}}_n(y) = h^d \widehat{p}_h(x). \quad (18)$$

Also note

$$\widehat{\mathbb{P}}^*_n(f_x) = \int K\left(\frac{x-y}{h}\right) d\widehat{\mathbb{P}}^*_n(y) = h^d \widehat{p}^*_h(x). \quad (19)$$

Therefore, for the bootstrap empirical process $\mathbb{G}^*_n = \sqrt{n}(\widehat{\mathbb{P}}^* - \widehat{\mathbb{P}})$,

$$\widehat{p}^*_h(x) - \widehat{p}_h(x) = \frac{1}{\sqrt{nh^d}} \mathbb{G}^*_n(f_x). \quad (20)$$

Thus, if we sample from $\widehat{p}_h$ and consider estimating $\widehat{p}_h$ by $\widehat{p}^*_h$, we are doing exactly the same procedure of estimating p_h by $\widehat{p}_h$. Therefore, Lemma 2 and Theorem 3 hold for approximating $\mathsf{Haus}(\widehat{D}_h^*, \widehat{D}_h)$ by a maxima for a Gaussian process. The difference is that the Gaussian process is defined on

$$\mathcal{F}_n = \left\{ f_x(y) \equiv \frac{1}{\sqrt{nh^d}\|\nabla \widehat{p}_h(x)\|} K\left(\frac{x-y}{h}\right) : x \in \widehat{D}_h \right\} \tag{21}$$

since the "parameter (level sets)" being estimated is $\widehat{D}_h$ (the estimator is $\widehat{D}_h^*$). Note that $\mathcal{F}_n$ is very similar to $\mathcal{F}$ except the denominator is slightly different and the support $\widehat{D}_h$ is also different from D_h. That is, we have

$$\begin{aligned} \sup_t \Bigg| \mathbb{P}\left(\sqrt{nh^d}\mathsf{Haus}(\widehat{D}_h^*, \widehat{D}_h) < t \middle| X_1, \cdots, X_n\right) \\ - \mathbb{P}\left(\sup_{f\in\mathcal{F}_n} |\mathbb{B}_n(f)| < t \middle| X_1, \cdots, X_n\right) \Bigg| \le O\left(\left(\frac{\log^7 n}{nh^d}\right)^{1/8}\right), \end{aligned} \tag{22}$$

where $\mathbb{B}_n$ is a Gaussian process on $\mathcal{F}_n$ such that for any $f_1, f_2 \in \mathcal{F}_n$,

$$\mathbb{E}(\mathbb{B}_n(f_1)|X_1, \cdots, X_n) = 0, \quad \mathrm{Cov}(\mathbb{B}_n(f_1), \mathbb{B}(f_2)|X_1, \cdots, X_n) = \frac{1}{n}\sum_{i=1}^n f_1(X_i) f_2(X_i). \tag{23}$$

Step 2: Support approximation. In this step, we will show that

$$\sup_{f\in\mathcal{F}_n} |\mathbb{B}_n(f)| \approx \sup_{f\in\mathcal{F}} |\mathbb{B}_n(f)| \approx \sup_{f\in\mathcal{F}} |\mathbb{B}(f)|. \tag{24}$$

The first approximation can be shown by using the Gaussian comparison lemma (Theorem 2 in Chernozhukov et al. (2014b); also see Lemma 17 in Chen et al. (2014b)). We do the same thing as Step 3 in the proof of Theorem 8 in Chen et al. (2014b) so we omit the details. Essentially, given any $\epsilon > 0$, we can construct a pair of balanced ϵ-nets for both $\mathcal{F}$ and $\mathcal{F}_n$, denoted as $\{g_1, \cdots, g_K\}$ and $\{g_1^n, \cdots, g_K^n\}$ so that $\max_j \|g_j - g_j^n\|_{\max}^* = O(\|\widehat{p}_h - p_h\|_{1,\max}^*)$. Then this ϵ-net leads to

$$\begin{aligned} \sup_t \Bigg| \mathbb{P}\left(\sup_{f\in\mathcal{F}_n} |\mathbb{B}_n(f)| < t \middle| X_1, \cdots, X_n\right) \\ - \mathbb{P}\left(\sup_{f\in\mathcal{F}} |\mathbb{B}_n(f)| < t \middle| X_1, \cdots, X_n\right) \Bigg| \le O\left(\left(\|\widehat{p}_h - p_h\|_{1,\max}^*\right)^{1/3}\right). \end{aligned} \tag{25}$$

The difference between $\sup_{f\in\mathcal{F}} |\mathbb{B}_n(f)|$ and $\sup_{f\in\mathcal{F}} |\mathbb{B}(f)|$ is small since the these two

Gaussian processes differ in their covariance but as $n \to \infty$, the covariances converges at rate $1/\sqrt{n}$ so that we can neglect the difference between them. Thus, combining (22) and (25) and the argument from previous paragraph, we conclude

$$\begin{aligned}\sup_t \left| \mathbb{P}\left(\sqrt{nh^d} \mathsf{Haus}(\widehat{D}_h^*, \widehat{D}_h) < t \middle| X_1, \cdots, X_n \right) - \mathbb{P}\left(\sup_{f \in \mathcal{F}} |\mathbb{B}(f)| < t \right) \right| \\ \leq O\left(\left(\frac{\log^7 n}{nh^d} \right)^{1/8} \right) + O\left(\left(\|\widehat{p}_h - p_h\|_{1,\max}^* \right)^{1/3} \right).\end{aligned} \tag{26}$$

Now comparing the above result to Theorem 3 and using the fact that the first big-O term dominates the second term (the first is of rate $-1/8$ for n but the second term is at rate $-1/6$ by Theorem 9), we conclude the result for first assertion.

For the coverage, let $W_n = \mathsf{Haus}(\widehat{D}_h, D_h)$ and $w_{n,1-\alpha} = F_{W_n}^{-1}(1-\alpha)$. Since $D_h \subset \widehat{D}_h \oplus \mathsf{Haus}(\widehat{D}_h, D_h)$, we have

$$\mathbb{P}(D_h \subset \widehat{D}_h \oplus w_{n,1-\alpha}) = 1 - \alpha. \tag{27}$$

Now by the first assertion, the difference for $w_{n,1-\alpha}$ and the bootstrap estimate $w_{n,1-\alpha}^*$ differs at rate $O\left(\left(\frac{\log^7 n}{nh^d} \right)^{1/8} \right)$, which completes the proof.

□

References

F. Chazal, A. Lieutier, and J. Rossignac. Normal-map between normal-compatible manifolds. *International Journal of Computational Geometry & Applications*, 17 (05):403–421, 2007.

Y.-C. Chen, C. R. Genovese, R. J. Tibshirani, and L. Wasserman. Nonparametric modal regression. *arXiv preprint arXiv:1412.1716*, 2014a.

Y.-C. Chen, C. R. Genovese, and L. Wasserman. Asymptotic theory for density ridges. *arXiv preprint arXiv:1406.5663*, 2014b.

V. Chernozhukov, D. Chetverikov, and K. Kato. Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics*, 42(5):1787–1818, 2014a.

V. Chernozhukov, D. Chetverikov, and K. Kato. Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, pages 1–24, 2014b.

V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, 42(4):1564–1597, 2014c.

A. Cuevas, W. González-Manteiga, and A. Rodríguez-Casal. Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19, 2006.

A. W. Van Der Vaart and J. A. Wellner. *Weak Convergence*. Springer, New York, 1996.

—