OPERATIONS RESEARCH
DOI 10.1287/opre.1100.0825ec

e-companion

ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"A Computable Plug-in Estimator of Minimum Volume Sets for Novelty Detection" by Chiwoo Park, Jianhua Z. Huang, and Yu Ding, *Operations Research*, DOI 10.1287/opre.1100.0825.

Material for on-line supplement

This e-companion provides a proof of consistency of the computable plug-in estimator, denoted as $MVC(\alpha;\hat{f}_n,P_n)$, which is a level set whose level is given by the solution of the following optimization problem:

$$\max\{y\in\mathbb{R}^+ : P_n(\hat{A}_{n,y})\geq\alpha\},\ \text{where } \hat{A}_{n,y}=\{x:\hat{f}_n(x)\geq y\}, \tag{EC.1}$$

where $P_n(A)=\frac{1}{n}\sum_{i=1}^n 1_A(x_i)$ is the empirical distribution for giving data points $x_1,..,x_n$ and $\hat{f}_n(x)$ is a kernel density estimator.

The strategy for the proof is to show that the volume and probability mass of the computable plug-in estimator are asymptotically equivalent to those of the original plug-in estimator $MVC(\alpha;\hat{f}_n)$, which is a level set whose level is the solution of the following optimization problem:

$$\max\{y\in\mathbb{R}^+ : \int_{\hat{A}_{n,y}}\hat{f}_n(x)\,dx\geq\alpha\},\ \text{where } \hat{A}_{n,y}=\{x:\hat{f}_n(x)\geq y\}. \tag{EC.2}$$

We give the proof of the consistency result after restating the the assumptions and the theorem. Recall that the minimum volume cut $MVC(\alpha;f)=\{x:f(x)\geq y^*\}$, where y^* solves the optimization problem

$$\max\{y\in\mathbb{R}^+ : \int_{A_y} f(x)\,dx\geq\alpha\},\ \text{where } A_y=\{x:f(x)\geq y\}. \tag{EC.3}$$

Let $\Theta\subset(0,\sup f)$ be an open interval that contains y^* and let $\|\cdot\|$ stand for the Euclidean norm over any finite-dimensional space. Let $A\Delta B=(A\cap B^c)\cup(A^c\cap B)$ denote the symmetric difference of sets A and B.

Assumption EC.1. *The kernel function K is continuously differentiable and has compact support. Moreover, there exists a monotone nondecreasing function $\mu:\mathbb{R}_+\to\mathbb{R}$ such that $K(x)=\mu(\|x\|)$ for all $x\in\mathbb{R}^d$.*

Assumption EC.2. *The density function f is twice continuously differentiable and $f(x)\to 0$ as $\|x\|\to\infty$.*

Assumption EC.3. *For any* $t \in \Theta$, $\inf_{f^{-1}(\{t\})} \|\nabla f\| > 0$, *where* $\nabla f(x)$ *is the gradient of* f *at* x.

Theorem EC.1. *Suppose that Assumptions EC.1, EC.2 and EC.3 hold. If the bandwidth* h_n *used in the kernel density estimation satisfies that* $nh_n^{d+4}(\log n)^2 \to 0$ *and* $nh_n^{d+2}/(\log n) \to \infty$, *then*

$$\int_{MVC(\alpha;\hat{f}_n,P_n)} f(x)\,dx \to \alpha \qquad \text{in probability.}$$

$$\lambda\{MVC(\alpha;\hat{f}_n,P_n)\,\Delta\,MVC(\alpha;f)\} \to 0 \qquad \text{in probability.}$$

Appendix A: Proof of Theorem EC.1

We show that the volume and probability mass of the computable plug-in estimator are asymptotically equivalent to those of the original plug-in estimator $MVC(\alpha;\hat{f}_n)$. Then, the consistency of the original estimator established in (Cadre 2006) will imply the consistency of the computable plug-in estimator.

Let $M = \sup_x f(x)$ and let $A_{n,k} = \{f \geq \frac{k}{n}\}$ for $k \in K_n := \{0,1,\dots,nM\}$. Here and throughout this appendix, we use the abbreviation $\{f \geq \frac{k}{n}\}$ to denote the set $\{x : f(x) \geq \frac{k}{n}\}$. For each positive integer n, define a class of indicator functions $G_n := \{1_{A_{n,0}}, 1_{A_{n,1}}, \cdots, 1_{A_{n,nM}}\}$. For $g \in G_n$, define

$$P_n(g) = \frac{1}{n}\sum_{i=1}^{n} g(x_i) \quad \text{and} \quad P(g) = E[g(x_1)],$$

where $x_1, x_2, .., x_n$ are i.i.d. draws from the probability density f.

By Hoeffding's inequality (Hoeffding 1963), for any $\epsilon > 0$,

$$P(|P_n(g) - P(g)| \geq \epsilon) \leq 2\exp(-2n\epsilon^2), \qquad g \in G_n. \tag{EC.4}$$

It follows that

$$P(\sup_{g\in G_n}\{|P_n(g) - P(g)| \geq \epsilon\}) \leq \sum_{g\in G_n} P(|P_n(g) - P(g)| \geq \epsilon) \leq 2nM\exp\{-2n\epsilon^2\}. \tag{EC.5}$$

Thus,

$$\sum_{n=1}^{\infty} P(\sup_{g\in G_n}\{|P_n(g) - P(g)| \geq \epsilon\}) \leq \sum_{n=1}^{\infty} 2nM\exp\{-2n\epsilon^2\} < \infty.$$

By the reverse Fatou's Lemma and the Borel-Cantelli Lemma (Williams 1991), there exists a $L > 0$ such that

$$\sup_{n\geq L} P(\sup_{g\in G_n}\{|P(g)-P_n(g)|\geq\epsilon\}) \leq P\left(\bigcup_{n\geq L}\sup_{g\in G_n}\{|P(g)-P_n(g)|\geq\epsilon\}\right)\leq\epsilon. \tag{EC.6}$$

Let $z_n > 0$ be a solution of optimization problem (EC.1) and let $\hat{A}_{n,z_n} = \{\hat{f}_n \geq z_n\}$. According to Assumption EC.1, Assumption EC.2 and Pollard (1984, Example 38 and Problem 28),

$$\lim_{n\to\infty}\sup_x |\hat{f}_n(x) - f(x)| = 0, \qquad \text{almost surely.} \tag{EC.7}$$

Thus, there exists a large integer N_1 such that for $n \geq N_1$,

$$\sup_x |\hat{f}_n(x) - f(x)| \leq \epsilon, \qquad \text{almost surely.} \tag{EC.8}$$

Thus, $\{f \geq z_n + \epsilon\} \subset \{\hat{f}_n \geq z_n\} \subset \{f \geq z_n - \epsilon\}$ almost surely for $n \geq N_1$. Let $N_2 \geq \max(L, N_1)$. We can choose k $(0 \leq k \leq N_2 M)$ such that $\frac{k+1}{N_2} \geq z_{N_2} + \epsilon$ and $\frac{k}{N_2} \leq z_{N_2} - \epsilon$. Then, almost surely, the following holds:

$$\{f \geq \frac{k+1}{N_2}\} \subset \{\hat{f}_{N_2} \geq z_{N_2}\} \subset \{f \geq \frac{k}{N_2}\}$$

It follows that

$$\lambda(\{f \geq \frac{k}{N_2}\}) - \lambda(\{\hat{f}_{N_2} \geq z_{N_2}\}) \leq \lambda(\{f \geq \frac{k}{N_2}\}) - \lambda(\{f \geq \frac{k+1}{N_2}\}). \tag{EC.9}$$

Because of Assumptions EC.2 and EC.3, by Cadre (2006, Proposition A.2),

$$\lambda(\{f \geq \frac{k}{N_2}\}) - \lambda(\{f \geq \frac{k+1}{N_2}\}) \leq \epsilon. \tag{EC.10}$$

The inequalities (EC.9) and (EC.10) together imply that

$$\lambda(\{f \geq \frac{k}{N_2}\}) - \lambda(\{\hat{f}_{N_2} \geq z_{N_2}\}) \leq \epsilon. \tag{EC.11}$$

Define $H = \bigcap_{g\in G_{N_2}}\{|P(g) - P_{N_2}(g)| < \epsilon\}$. Then, by inequality (EC.6),

$$\begin{aligned} P(H) = 1 - P(H^c) &= 1 - P\left(\bigcup_{g\in G_{N_2}}\{|P(g)-P_{N_2}(g)|\geq\epsilon\}\right) \\ &\geq 1 - \sup_{n\geq L} P\left(\bigcup_{g\in G_n}\{|P(g)-P_n(g)|\geq\epsilon\}\right) \geq 1-\epsilon. \end{aligned} \tag{EC.12}$$

Let $g_1 = 1_{A_{N_2,k}}$ and $g_2 = 1_{A_{N_2,k+1}}$. Since g_1 and g_2 are in G_{N_2}, by (EC.12), $|P(g_1) - P_{N_2}(g_1)| < \epsilon$ and $|P(g_2) - P_{N_2}(g_2)| < \epsilon$ with probability at least $1-\epsilon$. Using this result, the triangle inequality, and (EC.10), we obtain that, with probability at least $1-\epsilon$,

$$\begin{aligned}
|P_{N_2}(g_1) - P_{N_2}(g_2)| &= \left| \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\{f \ge \frac{k}{N_2}\}}(x_i) - \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\{f \ge \frac{k+1}{N_2}\}}(x_i) \right| \\
&\le \left| \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\{f \ge \frac{k}{N_2}\}}(x_i) - \int_{f \ge \frac{k}{N_2}} f \right| \\
&\quad + \left| \int_{f \ge \frac{k+1}{N_2}} f - \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\{f \ge \frac{k+1}{N_2}\}}(x_i) \right| + \left| \int_{f \ge \frac{k}{N_2}} f - \int_{f \ge \frac{k+1}{N_2}} f \right| \qquad \text{(EC.13)} \\
&< |P(g_1) - P_{N_2}(g_1)| \\
&\quad + |P_{N_2}(g_2) - P(g_2)| + M\left[\lambda(\{f \ge \frac{k}{N_2}\}) - \lambda(\{f \ge \frac{k+1}{N_2}\})\right] \\
&\le (M+2)\epsilon.
\end{aligned}$$

Applying the results of (EC.11), (EC.6) and (EC.13) in order, we obtain that, with at least probability $1-\epsilon$,

$$\begin{aligned}
\left| P(1_{\hat{f}_{N_2} \ge z_{N_2}}) - P_{N_2}(1_{\hat{f}_{N_2} \ge z_{N_2}}) \right| &= \left| \int_{\{\hat{f}_{N_2} \ge z_{N_2}\}} f - \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\{\hat{f}_{N_2} \ge z_{N_2}\}}(x_i) \right| \\
&\le \left| \int_{\{\hat{f}_{N_2} \ge z_{N_2}\}} f - \int_{\{f \ge \frac{k}{N_2}\}} f \right| \\
&\quad + \left| \int_{\{f \ge \frac{k}{N_2}\}} f - \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\{f \ge \frac{k}{N_2}\}}(x_i) \right| \\
&\quad + \left| \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\{f \ge \frac{k}{N_2}\}}(x_i) - \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\{\hat{f}_{N_2} \ge z_{N_2}\}}(x_i) \right| \\
&< M\epsilon + \epsilon + (M+2)\epsilon = (2M+3)\epsilon.
\end{aligned}$$

That is, the following holds

$$P\left(\left| \int_{\hat{A}_{N_2,z_{N_2}}} f - \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\hat{A}_{N_2,z_{N_2}}}(x_i) \right| < (2M+3)\epsilon \right) \ge 1-\epsilon. \qquad \text{(EC.14)}$$

By the definition of $\hat{A}_{N_2,z_{N_2}}$,

$$\alpha \le \frac{1}{N_2}\sum_{i=1}^{N_2} 1_{\hat{A}_{N_2,z_{N_2}}}(x_i) \le \alpha + \frac{1}{N_2}. \qquad \text{(EC.15)}$$

By the results of (EC.14) and (EC.15),

$$P\left(\left|\int_{\hat{A}_{N_2,z_{N_2}}} f-\alpha\right| \le \epsilon+\frac{1}{N_2}\right) \ge 1-\epsilon.$$

Let $N_2 \to \infty$ and $\epsilon \to 0$ to get the following result:

$$\int_{\hat{A}_{n,z_n}} f(x) \to \alpha \quad \text{in probability.} \tag{EC.16}$$

The proof of the first part of the theorem is complete.

Let $\epsilon_n = \sup_x |\hat{f}_n(x) - f(x)|$. According to the uniform convergence (EC.7), $\epsilon_n \to 0$ almost surely. Observe that

$$\begin{aligned}\left|\int_{f\ge z_n} f - \int_{\hat{A}_{n,z_n}} f\right| &\le \int f\,|1_{\{f\ge z_n\}} - 1_{\{\hat{f}_n\ge z_n\}}| \\ &\le \int f\,1_{\{z_n-\epsilon_n\le f\le z_n-\epsilon_n\}} \\ &\le M\lambda(\{z_n-\epsilon_n \le f \le z_n-\epsilon_n\}\cap(0,\sup f]),\end{aligned} \tag{EC.17}$$

which tends to 0 as $n\to\infty$ according to Cadre (2006, Proposition A.2). This together with (EC.16) imply that

$$\int_{f\ge z_n} f \to \int_{f\ge y^*} f, \tag{EC.18}$$

where the level y^* satisfies $\int_{f\ge y^*} f = \alpha$. According to Cadre (2006, Proposition A.2), the map $s \mapsto \int_{f\ge s} f$ is continuous and one-to-one, so $z_n \to y^*$.

Let y_n be the solution of optimization problem (EC.2), which defines the original plug-in estimator. According to Cadre (2006, Corollary 2.1),

$$\int_{\hat{A}_{n,y_n}} f \to \int_{f\ge y^*} f = \alpha \qquad \text{in probability.} \tag{EC.19}$$

This together with (EC.16) and the triangle inequality yields

$$\int_{\hat{A}_{n,z_n}\Delta\hat{A}_{n,y_n}} f \le \left|\int_{\hat{A}_{n,z_n}} f-\alpha\right| + \left|\int_{\hat{A}_{n,y_n}} f-\alpha\right| \to 0 \quad \text{in probability.}$$

Since $\int_{\hat{A}_{n,z_n}\Delta\hat{A}_{n,y_n}} f \ge (\min\{y_n,z_n\}-\epsilon_n)\lambda(\hat{A}_{n,z_n}\Delta\hat{A}_{n,y_n})$, we obtain

$$(\min\{y_n,z_n\}-\epsilon_n)\lambda(\hat{A}_{n,z_n}\Delta\hat{A}_{n,y_n}) \to 0 \quad \text{in probability.}$$

Because $z_n \to y^*$, $y_n \to y^*$ (Cadre 2006, Lemma 4.3) and $\epsilon_n \to 0$, the above result implies that $\lambda(\hat{A}_{n,z_n} \Delta \hat{A}_{n,y_n}) \to 0$ in probability. It follows from Cadre (2006, Corollary 2.1) that $\lambda(\hat{A}_{n,y_n} \Delta\, MVC(\alpha; f)) \to 0$ in probability. Therefore, $\lambda(\hat{A}_{n,z_n} \Delta\, MVC(\alpha; f)) \to 0$ in probability.

References

Cadre, B. 2006. Kernel estimation of density level sets. *Journal of Multivariate Analysis* **97**(4) 999–1023.

Hoeffding, W. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**(301) 13–30.

Pollard, D. 1984. *Convergence of Stochastic Processes*. Springer, New York, NY, USA.

Williams, D. 1991. *Probability with Martingales*. Cambridge University Press, Cambridge, UK.