12 Limiting distribution of M-estimators

Let $X_1, \ldots, X_n$ be i.i.d. P observations taking values in a space $\mathcal{X}$. Let Θ denote a parameter space (assumed to be a metric space with metric $d(\cdot, \cdot)$) and, for each $\theta \in \Theta$, let m_θ denote a real-valued function on $\mathcal{X}$. Consider the map

$$\theta \mapsto \mathbb{M}_n(\theta) := \mathbb{P}_n[m_\theta(X)] \equiv \frac{1}{n}\sum_{i=1}^{n} m_\theta(X_i)$$

and let $\hat{\theta}_n$ denote the *maximizer* of $\mathbb{M}_n(\theta)$ over $\theta \in \Theta$, i.e.,

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} \mathbb{M}_n(\theta).$$

Such a quantity $\hat{\theta}_n$ is called an *M-estimator*. We study the (limiting) distribution of M-estimators (properly standardized) in this section.

Their statistical properties of $\hat{\theta}_n$ depend crucially on the behavior of the criterion function $\mathbb{M}_n(\theta)$ as $n \to \infty$. For example, we may ask: is $\hat{\theta}_n$ converging to some $\theta_0 \in \Theta$, as $n \to \infty$? A natural way to tackle the question is as follows: We expect that for each $\theta \in \Theta$, $\mathbb{M}_n(\theta)$ will be close to its population version

$$M(\theta) := P[m_\theta(X)], \qquad \theta \in \Theta.$$

Let

$$\theta_0 := \arg\max_{\theta \in \Theta} M(\theta).$$

If $\mathbb{M}_n$ and M are uniformly close, then maybe their argmax's $\hat{\theta}_n$ and θ_0 are also close. A key tool to studying such behavior of $\hat{\theta}_n$ is the *argmax continuous mapping theorem* which we consider next. Before we present the result in a general setup let us discuss the main idea behind the proof. For any given $\epsilon > 0$, we have to bound the probability $\mathbb{P}(d(\hat{\theta}_n, \theta_0) \geq \epsilon)$. The key step is to realize that

$$\begin{aligned} \mathbb{P}(d(\hat{\theta}_n, \theta_0) \geq \epsilon) &\leq \mathbb{P}\left(\sup_{\theta \in \Theta : d(\theta, \theta_0) \geq \epsilon} [\mathbb{M}_n(\theta) - \mathbb{M}_n(\theta_0)] > 0 \right) \\ &\leq \mathbb{P}\left(\sup_{\theta \in \Theta : d(\theta, \theta_0) \geq \epsilon} [(\mathbb{M}_n - M)(\theta) - (\mathbb{M}_n - M)(\theta_0)] > - \sup_{d(\theta, \theta_0) \geq \epsilon} [M(\theta) - M(\theta_0)] \right). \end{aligned} \tag{163}$$

The (uniform) closeness of $\mathbb{M}_n$ and M (cf. condition (3) in Theorem 12.1 below) shows that the left-hand side of (163) must converge to 0 (in probability), whereas if M has a *well-separated* unique maximum[120] (cf. condition (1) in Theorem 12.1) then the right-hand side of (163) must exceed a positive number, thereby showing that $\mathbb{P}(d(\hat{\theta}_n, \theta_0) \geq \epsilon) \to 0$ as $n \to \infty$. This was carried out in Subsection 3.5.1 while discussing the consistency of M-estimators.

[120] i.e., the function $M(\theta)$ should be strictly smaller than $M(\theta_0)$ on the complement of every neighborhood of the point θ_0.

12.1 Argmax continuous mapping theorems

We state our first argmax continuous mapping theorem below which generalizes the above discussed setup (so that it can also be used to derive asymptotic distributions of the M-estimator). Our first result essentially says that the argmax functional is *continuous* at functions M that have a well-separated unique maximum.

Theorem 12.1. *Let H be a metric space and let $\{\mathbb{M}_n(h), h \in H\}$ and $\{M(h), h \in H\}$ be stochastic processes indexed by H. Suppose the following conditions hold:*

1. *$\hat{h}$ is a random element of H which satisfies*

$$M(\hat{h}) > \sup_{h \notin G} M(h) \qquad a.s.,$$

for every open set G containing $\hat{h}$; i.e., M has a unique "well-separated" point of maximum.

2. *For each n, let $\hat{h}_n \in H$ satisfy*

$$\mathbb{M}_n(\hat{h}_n) \geq \sup_{h \in H} \mathbb{M}_n(h) - o_{\mathbb{P}}(1).$$

3. *$\mathbb{M}_n \xrightarrow{d} M$ in $\ell^\infty(H)$.*

Then $\hat{h}_n \xrightarrow{d} \hat{h}$ in H.

Proof. By the Portmanteau theorem 10.5, to prove $\hat{h}_n \xrightarrow{d} \hat{h}$ it suffices to show that

$$\limsup_{n \to \infty} \mathbb{P}^*\{\hat{h}_n \in F\} \leq \mathbb{P}\{\hat{h} \in F\} \tag{164}$$

for every closed subset F of H. Fix a closed set F and note that

$$\{\hat{h}_n \in F\} \subseteq \left\{ \sup_{h \in F} \mathbb{M}_n(h) \geq \sup_{h \in H} \mathbb{M}_n(h) - o_{\mathbb{P}}(1) \right\}.$$

Therefore,

$$\mathbb{P}^*\big(\hat{h}_n \in F\big) \leq \mathbb{P}^*\Big(\sup_{h \in F} \mathbb{M}_n(h) - \sup_{h \in H} \mathbb{M}_n(h) + o_{\mathbb{P}}(1) \geq 0 \Big).$$

The map $\sup_{h \in F} \mathbb{M}_n(h) - \sup_{h \in H} \mathbb{M}_n(h)$ converges in distribution to $\sup_{h \in F} M(h) - \sup_{h \in H} M(h)$ as $\mathbb{M}_n \xrightarrow{d} M$ in $\ell^\infty(H)$ and by the continuous mapping theorem. We thus have

$$\limsup_{n \to \infty} \mathbb{P}^*\Big(\hat{h}_n \in F\Big) \leq \mathbb{P}\left(\sup_{h \in F} M(h) \geq \sup_{h \in H} M(h) \right),$$

where we have again used the Portmanteau theorem. The first assumption of the theorem implies that $\{\sup_{h \in F} M(h) \geq \sup_{h \in H} M(h)\} \subseteq \{\hat{h} \in F\}$ (note that F^c is open). This proves (164). □

The idea behind the proof of the above theorem can be used to prove the following stronger technical lemma.

Lemma 12.2. *Let H be a metric space and let $\{\mathbb{M}_n(h) : h \in H\}$ and $\{M(h) : h \in H\}$ be stochastic processes indexed by H. Let A and B be arbitrary subsets of H. Suppose the following conditions hold:*

1. *$\hat{h}$ is a random element of H which satisfies $M(\hat{h}) > \sup_{h \in A \cap G^c} M(h)$ almost surely for every open set G containing $\hat{h}$.*
2. *For each n, let $\hat{h}_n \in H$ be such that $\mathbb{M}_n(\hat{h}_n) \geq \sup_{h \in H} \mathbb{M}_n(h) - o_{\mathbb{P}}(1)$.*
3. *$\mathbb{M}_n \xrightarrow{d} M$ in $\ell^\infty(A \cup B)$.*

Then

$$\limsup_{n \to \infty} \mathbb{P}^*\left(\hat{h}_n \in F \cap A\right) \leq \mathbb{P}\left(\hat{h} \in F\right) + \mathbb{P}\left(\hat{h} \in B^c\right) \tag{165}$$

for every closed set F.

Observe that Theorem 12.1 is a special case of this lemma which corresponds to $A = B = H$.

Proof of Lemma 12.2. The proof is very similar to that of Theorem 12.1. Observe first that

$$\left\{\hat{h}_n \in F \cap A\right\} \subseteq \left\{\sup_{h \in F \cap A} \mathbb{M}_n(h) - \sup_{h \in B} \mathbb{M}_n(h) + o_P(1) \geq 0\right\}.$$

The term $\sup_{h \in F \cap A} \mathbb{M}_n(h) - \sup_{h \in B} \mathbb{M}_n(h) + o_P(1)$ converges in distribution to $\sup_{h \in F \cap A} M(h) - \sup_{h \in B} M(h)$ because $\mathbb{M}_n \xrightarrow{d} M$ in $\ell^\infty(A \cup B)$. This therefore gives

$$\limsup_{n \to \infty} \mathbb{P}^*\left(\hat{h}_n \in F \cap A\right) \leq \mathbb{P}\left(\sup_{h \in F \cap A} M(h) - \sup_{h \in B} M(h) \geq 0\right)$$

Now if the event $\{\sup_{h \in F \cap A} M(h) \geq \sup_{h \in B} M(h)\}$ holds and if $\hat{h} \in B$, then $\sup_{h \in F \cap A} M(h) \geq M(\hat{h})$ which can only happen if $\hat{h} \in F$. This means

$$\mathbb{P}\left(\sup_{h \in F \cap A} M(h) - \sup_{h \in B} M(h) \geq 0\right) \leq \mathbb{P}(\hat{h} \in B^c) + \mathbb{P}(\hat{h} \in F)$$

which completes the proof. □

We next prove a more applicable argmax continuous mapping theorem. The assumption that $\mathbb{M}_n \xrightarrow{d} M$ in $\ell^\infty(H)$ is too stringent. It is much more reasonable to assume that $\mathbb{M}_n \xrightarrow{d} M$ in $\ell^\infty(K)$ for every *compact* subset K of H. The next theorem proves that $\hat{h}_n$ converges in law to $\hat{h}$ under this weaker assumption.

As we will be restricting analysis to compact sets in the next theorem, we need to assume that $\hat{h}_n$ and $\hat{h}$ lie in compact sets with arbitrarily large probability. This condition, made precise below, will be referred to as the *tightness condition*:

For every $\epsilon > 0$, there exists a compact set $K_\epsilon \subseteq H$ such that

$$\limsup_{n\to\infty} \mathbb{P}^*\Big(\hat{h}_n \notin K_\epsilon\Big) \leq \epsilon \quad \text{and} \quad \mathbb{P}\Big(\hat{h} \notin K_\epsilon\Big) \leq \epsilon. \tag{166}$$

Theorem 12.3 (Argmax continuous mapping theorem)**.** *Let H be a metric space and let $\{\mathbb{M}_n(h) : h \in H\}$ and $\{M(h) : h \in H\}$ be stochastic processes indexed by H. Suppose that the following conditions hold:*

1. *$\mathbb{M}_n \xrightarrow{d} M$ in $\ell^\infty(K)$ for every compact subset K of H.*
2. *Almost all sample paths $h \mapsto M(h)$ are upper semicontinuous[121] (u.s.c.) and possess a unique maximum at a random point $\hat{h}$.*
3. *For each n, let $\hat{h}_n$ be a random element of H such that $\mathbb{M}_n(\hat{h}_n) \geq \sup_{h\in H} \mathbb{M}_n(h) - o_{\mathbb{P}}(1)$.*
4. *The tightness condition (166) holds.*

Then $\hat{h}_n \xrightarrow{d} \hat{h}$ in H.

Proof. Let K be an arbitrary compact subset of H. We first claim that

$$M(\hat{h}) > \sup_{h\in K\cap G^c} M(h)$$

for every open set G containing $\hat{h}$. Suppose, for the sake of contradiction, that $M(\hat{h}) = \sup_{h\in K\cap G^c} M(h)$ for some open set G containing $\hat{h}$. In that case, there exist $h_m \in K \cap G^c$ with $M(h_m) \to M(h)$ as $m \to \infty$. Because $K \cap G^c$ (intersection of a closed set with a compact set) is compact, a subsequence of $\{h_m\}$ converges which means that we can assume, without loss of generality, that $h_m \to h$ for some $h \in K \cap G^c$. By the u.s.c. hypothesis, this implies that $\limsup_{m\to\infty} M(h_m) \leq M(h)$ which is same as $M(\hat{h}) \leq M(h)$. This implies that $\hat{h}$ is not a unique maximum (as $\hat{h} \in G$ and $h \in G^c$, we note that $\hat{h} \neq h$). This proves the claim.

We now use Lemma 12.2 with $A = B = K$ (note that $\mathbb{M}_n \xrightarrow{d} M$ on $\ell^\infty(A\cup B) = \ell^\infty(K)$). This gives that for every closed set F, we have

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}^*\Big(\hat{h}_n \in F\Big) &\leq \limsup_{n\to\infty} \mathbb{P}^*\Big(\hat{h}_n \in F \cap K\Big) + \limsup_{n\to\infty} \mathbb{P}^*\Big(\hat{h}_n \in K^c\Big) \\ &\leq \mathbb{P}\Big(\hat{h} \in F\Big) + \mathbb{P}\Big(\hat{h} \in K^c\Big) + \limsup_{n\to\infty} \mathbb{P}^*\Big(\hat{h}_n \in K^c\Big).\end{aligned}$$

The term on the right hand side above can be made smaller than $\mathbb{P}\big(\hat{h} \in F\big) + \epsilon$ for every $\epsilon > 0$ by choosing K appropriately (using tightness). An application of the Portmanteau theorem now completes the proof. □

[121] Recall the definition of upper semicontinuity: f is u.s.c. at x_0 if $\limsup_{n\to\infty} f(x_n) \leq f(x_0)$ whenever $x_n \to x_0$ as $n \to \infty$.

As a simple consequence of Theorems 12.1 and 12.3, we can prove the following theorem which is useful for checking consistency of M-estimators. Note that $\mathbb{M}_n \xrightarrow{d} M$ for a deterministic process M is equivalent to $\mathbb{M}_n \xrightarrow{\mathbb{P}} M$. This latter statement is equivalent to $\sup_{h \in H} |\mathbb{M}_n(h) - M(h)|$ converges to 0 in probability.

Theorem 12.4 (Consistency Theorem)**.** *Let Θ be a metric space. For each $n \geq 1$, let $\{\mathbb{M}_n(\theta) : \theta \in \Theta\}$ be a stochastic process. Also let $\{M(\theta) : \theta \in \Theta\}$ be a deterministic process.*

1. *Suppose $\sup_{\theta \in \Theta} |\mathbb{M}_n(\theta) - M(\theta)| \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. Also suppose the existence of $\theta_0 \in \Theta$ such that $M(\theta_0) > \sup_{\theta \notin G} M(\theta)$ for every open set G containing θ_0. Then any sequence sequence of M-estimators $\hat{\theta}_n$ (assuming that $\mathbb{M}_n(\hat{\theta}_n) \geq \sup_{\theta \in \Theta} \mathbb{M}_n(\theta) - o_P(1)$ is enough), converges in probability to θ_0.*
2. *Suppose $\sup_{\theta \in K} |\mathbb{M}_n(\theta) - M(\theta)| \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$ for every compact subset K of Θ. Suppose also that the deterministic limit process M is upper semicontinuous and has a unique maximum at θ_0. If $\{\hat{\theta}_n\}$ is tight, then $\hat{\theta}_n$ converges to θ_0 in probability.*

Remark 12.1. *For M-estimators, we can apply the above theorem with $\mathbb{M}_n(\theta) := \sum_{i=1}^n m_\theta(X_i)/n$ and $M(\theta) := P[m_\theta]$. In this case, the condition $\sup_{\theta \in K} |\mathbb{M}_n(\theta) - M(\theta)| \xrightarrow{\mathbb{P}} 0$ is equivalent to $\{m_\theta : \theta \in K\}$ being P-Glivenko-Cantelli.*

Theorem 12.3 can also be used to prove asymptotic distribution results for M-estimators, as illustrated in the following examples.

12.2 Asymptotic distribution

In this section we present one result that gives the asymptotic distribution of M-estimators for the case of i.i.d. observations. The formulation is from [van der Vaart, 1998]. The limit distribution of the sequence $\sqrt{n}(\hat{\theta}_n - \theta_0)$ follows from the following theorem, where $\hat{\theta}_n$ is an M-estimator of the finite dimensional parameter θ_0 (i.e., $\hat{\theta}_n := \arg\max_{\theta \in \Theta} \mathbb{M}_n(\theta)$ where $\mathbb{M}_n(\theta) = \mathbb{P}_n[m_\theta(X)]$).

Example 12.5 (Parametric maximum likelihood estimators)**.** *Suppose $X_1, \ldots, X_n$ are i.i.d. from an unknown density p_{θ_0} belonging to a known class $\{p_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\}$. Let $\hat{\theta}_n$ denote the maximum likelihood estimator of θ_0. A classical result is that, under some smoothness assumptions, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution to $N_k(0, I^{-1}(\theta_0))$ where $I(\theta_0)$ denotes the Fisher information matrix.*

This result can be derived from the argmax continuous mapping theorem. The first step is to observe that if $\theta \mapsto p_\theta(x)$ is sufficiently smooth at θ_0, then, for any $h \in \mathbb{R}^k$,

$$\sum_{i=1}^n \log \frac{p_{\theta_0 + hn^{-1/2}}(X_i)}{p_{\theta_0}(X_i)} = h^\top \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\theta_0}(X_i) - \frac{1}{2} h^\top I(\theta_0) h + o_{P_{\theta_0}}(1) \tag{167}$$

where $\dot{\ell}_{\theta_0}(x) := \nabla_\theta \log p_\theta(x)$ *denotes the score function. Condition* (167) *is known as the LAN (local asymptotic normality) condition. We shall prove the asymptotic normality of* $\hat{\theta}_n$ *assuming the marginal convergence of* (167) *(for every fixed* h*) can be suitably strengthened to a process level result in* $\ell^\infty(K)$*, for* $K \subset \mathbb{R}^k$ *compact. We apply the argmax continuous mapping theorem (Theorem 12.3) with* $H = \mathbb{R}^k$*,*

$$\mathbb{M}_n(h) := \sum_{i=1}^n \log \frac{p_{\theta_0+hn^{-1/2}}(X_i)}{p_{\theta_0}(X_i)} \quad \text{and} \quad M(h) := h^T\Delta - \frac{1}{2}h^T I(\theta_0)h$$

where $\Delta \sim N_k(0, I(\theta_0))$*. Then* $\hat{h}_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ *and* $\hat{h} \sim N(0, I^{-1}(\theta_0))$*. The argmax theorem will then imply the result provided the conditions of the argmax theorem hold. The main condition is tightness of* $\{\hat{h}_n\}$ *which means that the rate of convergence of* $\hat{\theta}_n$ *to* θ_0 *is* $n^{-1/2}$*.*

The above idea can be easily extended to derive the asymptotic distributions of other $\sqrt{n}$-consistent estimators, e.g., non-linear regression, robust regression, etc. (see [van der Vaart, 1998, Chapter 5] for more details).

Theorem 12.6. *Suppose that* $x \mapsto m_\theta(x)$ *is a measurable function for each* $\theta \in \Theta \subset \mathbb{R}^d$ *for an open set* Θ*, that* $\theta \mapsto m_\theta(x)$ *is differentiable at* $\theta_0 \in \Theta$ *for* P*-almost every* x *with derivative* $\dot{m}_{\theta_0}(x)$*, and that*

$$|m_{\theta_1}(x) - m_{\theta_2}(x)| \le F(x)\|\theta_1 - \theta_2\| \tag{168}$$

holds for all θ_1, θ_2 *in a neighborhood of* θ_0*, where* $F \in L_2(P)$*. Also suppose that* $M(\theta) = P[m_\theta]$ *has a second order Taylor expansion*

$$P[m_\theta] - P[m_{\theta_0}] = \frac{1}{2}(\theta - \theta_0)^\top V(\theta - \theta_0) + o(\|\theta - \theta_0\|^2)$$

where θ_0 *is a point of maximum of* M *and* V *is symmetric and nonsingular (negative definite since* M *is a maximum at* θ_0*). If* $\mathbb{M}_n(\hat{\theta}_n) \ge \sup_\theta \mathbb{M}_n(\theta) - o_{\mathbb{P}}(n^{-1})$ *and* $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$*, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V^{-1}\mathbb{G}_n(\dot{m}_{\theta_0}) + o_{\mathbb{P}}(1) \xrightarrow{d} N_d(0, V^{-1}P[\dot{m}_{\theta_0}\dot{m}_{\theta_0}^\top]V^{-1}).$$

Proof. We will show that

$$\tilde{\mathbb{M}}_n(h) := n\mathbb{P}_n(m_{\theta_0+hn^{-1/2}} - m_{\theta_0}) \xrightarrow{d} h^\top\mathbb{G}(\dot{m}_{\theta_0}) + \frac{1}{2}h^\top Vh =: \mathbb{M}(h) \text{ in } \ell^\infty(\{h : \|h\| \le K\})$$

for every $K > 0$. Then the conclusion follows from the argmax continuous Theorem 12.3 upon noticing that

$$\hat{h} = \operatorname*{argmax}_h \mathbb{M}(h) = -B^{-1}\mathbb{G}(\dot{m}_{\theta_0}) \sim N_d(0, V^{-1}P(\dot{m}_{\theta_0}\dot{m}_{\theta_0}^\top)V^{-1}).$$

Now, observe that

$$n\mathbb{P}_n(m_{\theta_0+hn^{-1/2}} - m_{\theta_0}) = \sqrt{n}(\mathbb{P}_n - P)[\sqrt{n}(m_{\theta_0+hn^{-1/2}} - m_{\theta_0})] + nP(m_{\theta_0+hn^{-1/2}} - m_{\theta_0}).$$

By the second order Taylor expansion of $M(\theta) := P[m_\theta]$ about θ_0, the second term of the right side of the last display converges to $(1/2)h^\top V h$ uniformly for $\|h\| \le K$. To handle the first term we use the Donsker theorem with chaining classes. The classes

$$\mathcal{F}_n := \{\sqrt{n}(m_{\theta_0+hn^{-1/2}} - m_{\theta_0}) : \|h\| \le K\}$$

have envelopes $F_n = F = \dot{m}_{\theta_0}$ for all n, and since $\dot{m}_{\theta_0} \in L_2(P)$ the Lindeberg condition is satisfied easily. Furthermore, with

$$f_{n,g} = \sqrt{n}(m_{\theta_0+gn^{-1/2}} - m_{\theta_0}), \qquad f_{n,h} = \sqrt{n}(m_{\theta_0+hn^{-1/2}} - m_{\theta_0}),$$

by the dominated convergence theorem the covariance functions satisfy

$$P(f_{n,g}f_{n,h}) - P(f_{n,g})P(f_{n,h}) \to P(g^\top \dot{m}_{\theta_0}\dot{m}_{\theta_0}^\top h) = g^\top \mathbb{E}[\mathbb{G}(\dot{m}_{\theta_0})\mathbb{G}(\dot{m}_{\theta_0})^\top]h.$$

Finally, the bracketing entropy condition holds since, by way of the same entropy calculations used in the proof of we have

$$N_{[\,]}(2\epsilon\|F\|_{P,2}, \mathcal{F}_n, L_2(P)) \le \left(\frac{CK}{\epsilon}\right)^d, \quad \text{i.e.,} \quad N_{[\,]}(\epsilon, \mathcal{F}_n, L_2(P)) \lesssim \left(\frac{CK\|F\|_{P,2}}{\epsilon}\right)^d$$

Thus, $J_{[\,]}(\delta, \mathcal{F}_n, L_2(P)) \lesssim \int_0^\delta \sqrt{d \log\left(\frac{CK}{\epsilon}\right)}\, d\epsilon$, and hence the bracketing entropy hypothesis of Donsker theorem holds. We conclude that $\tilde{\mathbb{M}}_n(h)$ converges weakly to $h^\top \mathbb{G}(\dot{m}_{\theta_0})$ in $\ell^\infty(\{h : \|h\| \le K\})$, and the desired result holds. □

12.3 A non-standard example

Example 12.7 (Analysis of the shorth)**.** *Recall the setup of Example 5.4. Suppose that $X_1, \ldots, X_n$ are i.i.d. P on $\mathbb{R}$ with density p with respect to the Lebesgue measure. Let F_X be the distribution function of X. Suppose that p is a unimodal symmetric density with mode θ_0 (with $p'(x) > 0$ for $x < \theta_0$ and $p'(x) < 0$ for $x > \theta_0$). We want to estimate θ_0.*

Let

$$\mathbb{M}(\theta) := P[m_\theta] = \mathbb{P}(|X - \theta| \le 1) = F_X(\theta + 1) - F_X(\theta - 1)$$

where $m_\theta(x) = \mathbf{1}_{[\theta-1,\theta+1]}(x)$. We can how that $\theta_0 = \operatorname{argmax}_{\theta\in\mathbb{R}} \mathbb{M}(\theta)$.

We can estimate θ_0 by

$$\hat{\theta}_n := \operatorname*{argmax}_{\theta\in\mathbb{R}} \mathbb{M}_n(\theta), \qquad \text{where} \qquad \mathbb{M}_n(\theta) = \mathbb{P}_n m_\theta.$$

We have already seen that (in Example 5.4) $\tau_n := n^{1/3}(\hat{\theta}_n - \theta_0) = O_p(1)$. Let us here give a sketch of the limiting distribution of (the normalized version of) $\hat{\theta}_n$. Observe that

$$\tau_n = \operatorname*{argmax}_{h\in\mathbb{R}} \mathbb{M}_n(\theta_0 + hn^{-1/3}) = \operatorname*{argmax}_{h\in\mathbb{R}} n^{2/3}[\mathbb{M}_n(\theta_0 + hn^{-1/3}) - \mathbb{M}_n(\theta_0)].$$

The plan is to show that the localized (and properly normalized) stochastic process $\tilde{\mathbb{M}}_n(h) := n^{2/3}[\mathbb{M}_n(\theta_0+hn^{-1/3})-\mathbb{M}_n(\theta_0)]$ *converges in distribution to "something" so that we can apply the argmax continuous mapping theorem (Theorem 12.3) to deduce the limiting behavior of* τ_n*. Notice that,*

$$\begin{aligned}\tilde{\mathbb{M}}_n(h) &:= n^{2/3}\mathbb{P}_n[m_{\theta_0+hn^{-1/3}} - m_{\theta_0}] \\ &= n^{2/3}(\mathbb{P}_n - P)[m_{\theta_0+hn^{-1/3}} - m_{\theta_0}] + P[m_{\theta_0+hn^{-1/3}} - m_{\theta_0}],\end{aligned}$$

where the second term is

$$\begin{aligned}n^{2/3}[\mathbb{M}(\theta_0 + hn^{-1/3}) - \mathbb{M}(\theta_0)] &= n^{2/3}\mathbb{M}'(\theta_0)hn^{-1/3} + n^{2/3}\frac{1}{2}\mathbb{M}''(\theta^*)h^2n^{-2/3} \\ &\to \frac{1}{2}\mathbb{M}''(\theta_0)h^2 = \frac{1}{2}[p'(\theta_0+1) - p'(\theta_0-1)]h^2,\end{aligned}$$

uniformly in $|h| \le K$*, for any constant* K*. Note that as* $\mathbb{M}$ *is differentiable* $\mathbb{M}'(\theta_0) = p(\theta_0+1) - p(\theta_0-1) = 0$*. Thus, we want to study the empirical process* $\mathbb{G}_n$ *indexed by the collection of functions* $\mathcal{F}_n := \{n^{1/6}(m_{\theta_0+hn^{-1/3}} - m_{\theta_0}) : |h| \le K\}$*. Here we can apply a Donsker theorem for a family of functions depending on* n*, for example Theorem 11.7. Thus we need to check that (160) and (161) hold. Observe that*

$$\begin{aligned}&P[(f_{n,s} - f_{n,t})^2] - [P(f_{n,s} - f_{n,t})]^2 \\ &= n^{1/3}P[(m_{\theta_0+sn^{-1/3}} - m_{\theta_0+tn^{-1/3}})^2] - o(1) \\ &= n^{1/3}\Big\{P\mathbf{1}_{[\theta_0-1+sn^{-1/6},\theta_0-1+tn^{-1/6}]} + P\mathbf{1}_{[\theta_0+1+sn^{-1/6},\theta_0+1+tn^{-1/6}]}\Big\} + o(1) \qquad \text{if} \quad t > s \\ &\to [p(\theta_0-1) + p(\theta_0+1)]|s-t|.\end{aligned}$$

Thus, we can conclude that

$$n^{2/3}(\mathbb{P}_n - P)[m_{\theta_0+hn^{-1/3}} - m_{\theta_0}] \xrightarrow{d} a\mathcal{Z}(h)$$

where $a^2 := p(\theta_0+1) + p(\theta_0-1)$ *and* $\mathcal{Z}$ *is a standard two-sided Brownian motion process starting from 0 (show this!). We can now use the argmax continuous mapping theorem to conclude now that*

$$\tau_n = n^{1/3}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \operatorname*{argmax}_h [a\mathcal{Z}(h) - bh^2],$$

where $b := -\mathbb{M}''(\theta_0)/2$*.*