T. Christensen

1 Consistency of Extremum Estimators

Recall that an estimator $\hat{\theta}$ of θ_0 is *consistent* if

$$\hat{\theta} \to_p \theta_0 \quad \text{as } n \to \infty\,. \tag{1}$$

Consistency is a useful property. It says that as we observe more data, the probability of our estimator $\hat{\theta}$ being close to the estimand θ_0 should approach 1.

The following result is our master consistency result. The result can be used for M-estimation as well as GMM, SMM and MD. Let $\|\cdot\|$ denote a norm on Θ.

Theorem 1 (**Consistency of extremum estimators**). *Let the following hold:*
(i) (clean maximum) for any $\delta > 0$ *we have* $\sup_{\theta\in\Theta:\|\theta-\theta_0\|\geq\delta} Q(\theta) < Q(\theta_0)$
(ii) (uniform convergence) $\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| = o_p(1)$.
Then: any estimator $\hat{\theta}$ *that satisfies (2) is consistent, i.e.* $\hat{\theta} \to_p \theta$ *as* $n \to \infty$.

The intuition is as follows (also see Figure 1). The estimator $\hat{\theta}$ is obtained by maximizing Q_n:

$$Q_n(\hat{\theta}) > \sup_{\theta\in\Theta} Q_n(\theta) - \eta_n\,, \tag{2}$$

where $\eta_n \geq 0$ is $o_p(1)$. If θ_0 is identified, then the population objective function Q is uniquely maximized at θ_0. As $\hat{\theta}$ is obtained by maximizing Q_n and we know that Q_n becomes closer to Q as we observe more data, the maximum of Q_n should become closer to θ_0.

"Clean maximum" means $Q(\theta)$ can only approach $Q(\theta_0)$ as $\theta \to \theta_0$. This is needed to rule out situations in which $Q(\theta)$ may asymptote to $Q(\theta_0)$ as θ moves along certain directions (see Figure 2).

"Uniform convergence" means Q_n converges to Q in probability uniformly over the parameter space. This rules out, e.g., Q_n having a bump that moves around as n gets large.

For instance, suppose $\Theta = [-1, 1]$ and $Q : \Theta \to \mathbb{R}$ is continuous, with a unique maximum at $\theta_0 \neq 0$. Suppose also that

$$Q_n(\theta) = \left[\begin{array}{cl} Q(\theta) & \text{if } \theta \neq \frac{1}{n} \\ Q(\theta_0) + 1 & \text{if } \theta = \frac{1}{n} \end{array}\right. \tag{3}$$

Then $Q_n(\theta)$ converges pointwise to $Q(\theta)$ but not uniformly, because $\sup_\theta |Q_n(\theta) - Q(\theta)| \geq 1$. But also note that for each $n \geq 1$ the argmax of $Q_n(\theta)$ is $\hat{\theta} = \frac{1}{n}$, which converges to $0 \neq \theta_0$.

Proof of Theorem 1. We want to show that $\Pr(\|\hat{\theta} - \theta_0\| > \delta) \to 0$ (as $n \to \infty$) for each $\delta > 0$.

Fix any $\delta > 0$. Let $\epsilon = Q(\theta_0) - \sup_{\theta \in \Theta : \|\theta - \theta_0\| \geq \delta} Q(\theta)$. Note $\epsilon > 0$ by (i).

As $\eta_n = o_p(1)$ and $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| = o_p(1)$, we have with probability approaching one (wpa1) that

$$|\eta_n| < \frac{\epsilon}{3}, \quad \sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| < \frac{\epsilon}{3}. \tag{4}$$

Whenever these inequalities hold, we therefore have that

$$Q(\hat{\theta}) > Q(\theta_0) - \epsilon = \sup_{\theta \in \Theta : \|\theta - \theta_0\| \geq \delta} Q(\theta), \tag{5}$$

where the second equality is by definition of ϵ. It follows that $\|\hat{\theta} - \theta_0\| \leq \delta$ must hold whenever inequality (4) holds. But as (4) holds wpa1, we have therefore shown

$$\Pr(\|\hat{\theta} - \theta_0\| \leq \delta) \to 1, \tag{6}$$

as required. ∎

Remark 1. *If we assume $\eta_n = o_{a.s.}(1)$ and replace (ii) with $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| = o_{a.s.}(1)$, then we can show that $\hat{\theta} \to_{a.s.} \theta_0$ as $n \to \infty$. The proof is left as an exercise.*

2 Verifying Clean Maximum

There are many sufficient conditions for clean maximum. Here is one set:

Lemma 1 (Verifying "clean maximum"). *Let the following hold:*
(i) Θ is compact
(ii) $Q : \Theta \to \mathbb{R}$ is continuous
(iii) $Q(\theta_0) > Q(\theta)$ for each $\theta \in \Theta$ with $\theta \neq \theta_0$.
Then: "clean maximum" holds.

Proof. Fix any $\delta > 0$. The set $\{\theta \in \Theta : \|\theta - \theta_0\| \geq \delta\}$ is compact by (i). Then by (ii), we know that there is some $\theta^* \in \{\theta \in \Theta : \|\theta - \theta_0\| \geq \delta\}$ such that $\sup_{\theta \in \Theta : \|\theta - \theta_0\| \geq \delta} Q(\theta) = Q(\theta^*)$ and by (iii) we must have $Q(\theta^*) < Q(\theta_0)$. ∎

3 Verifying Uniform Convergence

This is done differently for M-estimators, GMM, SMM, and MD.

3.1 Consistency of M-Estimators

For M-estimators, it suffices to show that the following *uniform* law of large numbers holds:

$$\sup_{\theta \in \Theta} \left| \underbrace{\frac{1}{n} \sum_{t=1}^{n} m(X_t, \theta)}_{Q_n(\theta)} - \underbrace{\mathrm{E}[m(X_t, \theta)]}_{Q(\theta)} \right| = o_p(1)\,. \tag{7}$$

Note that this is a stronger notion than the law of large numbers which asserts the pointwise result

$$\frac{1}{n} \sum_{t=1}^{n} m(X_t, \theta) - \mathrm{E}[m(X_t, \theta) = o_p(1) \tag{8}$$

for each θ.

We establish uniform convergence using a notion of the "size" or "complexity" of the class of functions whose average we are taking. It will turn out that uniform convergence holds whenever the class of functions $\mathcal{M} = \{m(\cdot; \theta) : \theta \in \Theta\}$ is small enough that it has *finite bracketing numbers.* Later in the course, we will see that similar notions of size or complexity are used to establish convergence results for nonparametric and modern machine learning methods.

Let $L^1 = \{f(X_t) : \mathrm{E}[|f(X_t)|] < \infty\}$ and let $\mathcal{F} \subset L^1$ be a collection of functions of interest. The list of pairs of functions

$$l_{\varepsilon,1}, u_{\varepsilon,1}, l_{\varepsilon,2}, u_{\varepsilon,2}, \ldots, l_{\varepsilon,N}, u_{\varepsilon,N} \subset L^1 \tag{9}$$

is said to *bracket* $\mathcal{F}$ *at level* ε if for each $f \in \mathcal{F}$ we can choose a pair $l_{\varepsilon,i}$ and $u_{\varepsilon,i}$ such that $l_{\varepsilon,i} \leq f \leq u_{\varepsilon,i}$ and $\mathrm{E}[u_{\varepsilon,i} - l_{\varepsilon,i}] \leq \varepsilon$ for each i. The ε*-bracketing number* of $\mathcal{F}$, denoted $N_{[\,]}(\mathcal{F}, \varepsilon)$, is the minimal number pairs required to bracket $\mathcal{F}$ at level ε. If $N_{[\,]}(\mathcal{F}, \varepsilon) < \infty$ for all $\varepsilon > 0$ then we say that $\mathcal{F}$ *has finite bracketing numbers.*

Lemma 2 (**Uniform Strong Law of Large Numbers (ULLN)**). *Let the following hold:*
(i) $X_1, \ldots, X_n$ *are IID or SSE*
(ii) $N_{[\,]}(\mathcal{M}, \varepsilon) < \infty$ *for each* $\varepsilon > 0$.
Then:

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^{n} m(X_t, \theta) - \mathrm{E}[m(X_t, \theta)] \right| = o_{a.s.}(1)\,. \tag{10}$$

Proof. Take any rational $\varepsilon > 0$. For each $\theta \in \Theta$ there is a pair $l_{\varepsilon,i(\theta)}(X_t), u_{\varepsilon,i(\theta)}(X_t)$ with

$$l_{\varepsilon,i(\theta)}(X_t) \leq m(X_t;\theta) \leq u_{\varepsilon,i(\theta)}(X_t) \tag{11}$$

for all X_t and

$$\mathrm{E}[u_{\varepsilon,i(\theta)}(X_t) - l_{\varepsilon,i(\theta)}(X_t)] \leq \varepsilon \tag{12}$$

and $i(\theta) \in \{1, \ldots, N_{[\,]}(\mathcal{M}, \varepsilon)\}$. Therefore, for each $\theta \in \Theta$ we have

$$\begin{aligned}
Q_n(\theta) - Q(\theta) &= \frac{1}{n}\sum_{t=1}^{n} m(X_t;\theta) - \mathrm{E}[m(X_t;\theta)] \\
&\leq \frac{1}{n}\sum_{t=1}^{n} u_{\varepsilon,i(\theta)}(X_t) - \mathrm{E}[m(X_t;\theta)] && (13)\\
&= \frac{1}{n}\sum_{t=1}^{n} u_{\varepsilon,i(\theta)}(X_t) - \mathrm{E}[u_{\varepsilon,i(\theta)}(X_t)] + \big(\mathrm{E}[u_{\varepsilon,i(\theta)}(X_t)] - \mathrm{E}[m(X_t;\theta)]\big) && (14)\\
&\leq \frac{1}{n}\sum_{t=1}^{n} u_{\varepsilon,i(\theta)}(X_t) - \mathrm{E}[u_{\varepsilon,i(\theta)}(X_t)] + \varepsilon && (15)
\end{aligned}$$

because $\mathrm{E}[u_{\varepsilon,i(\theta)}(X_t)] - \mathrm{E}[m(X_t;\theta)] \leq \mathrm{E}[u_{\varepsilon,i(\theta)}(X_t)] - \mathrm{E}[l_{\varepsilon,i(\theta)}(X_t)] \leq \varepsilon$.

Taking the sup over $\theta \in \Theta$:

$$\begin{aligned}
\sup_{\theta\in\Theta} (Q_n(\theta) - Q(\theta)) &\leq \sup_{\theta\in\Theta}\left(\frac{1}{n}\sum_{t=1}^{n} u_{\varepsilon,i(\theta)}(X_t) - \mathrm{E}[u_{\varepsilon,i(\theta)}(X_t)]\right) + \varepsilon && (16)\\
&\leq \max_{1\leq i\leq N}\left(\frac{1}{n}\sum_{t=1}^{n} u_{\varepsilon,i}(X_t) - \mathrm{E}[u_{\varepsilon,i}(X_t)]\right) + \varepsilon && (17)
\end{aligned}$$

where $N = N_{[\,]}(\mathcal{M}, \varepsilon)$. Applying the SLLN or Ergodic Theorem yields

$$\frac{1}{n}\sum_{t=1}^{n} u_{\varepsilon,i}(X_t) - \mathrm{E}[u_{\varepsilon,i}(X_t)] \to_{a.s.} 0 \tag{18}$$

for each $1 \leq i \leq N_{[\,]}(\mathcal{M}, \varepsilon)$, and so

$$\max_{1\leq i\leq N}\left(\frac{1}{n}\sum_{t=1}^{n} u_{\varepsilon,i}(X_t) - \mathrm{E}[u_{\varepsilon,i}(X_t)]\right) \to_{a.s.} 0\,. \tag{19}$$

Therefore,

$$\sup_{\theta\in\Theta}(Q_n(\theta) - Q(\theta)) \leq \varepsilon + o_{a.s.}(1)\,. \tag{20}$$

A similar argument with the lower bracket gives us

$$\inf_{\theta\in\Theta}(Q_n(\theta)-Q(\theta)) \geq -\varepsilon + o_{a.s.}(1)\,. \tag{21}$$

Combining the preceding two inequalities, we obtain

$$\sup_{\theta\in\Theta}|Q_n(\theta)-Q(\theta)| \leq \varepsilon + o_{a.s.}(1)\,. \tag{22}$$

By definition of almost sure convergence, this means that there exists a set $S_\varepsilon \in \mathcal{F}$ with $\mathbb{P}(S_\varepsilon) = 1$ such that:

$$\limsup_{n\to\infty}\sup_{\theta\in\Theta}|Q_n(\theta;\omega)-Q(\theta)| \leq \varepsilon \tag{23}$$

for all $\omega \in S_\varepsilon$. Take $S = \cap_{\varepsilon\in\mathbb{Q}_+} S_\varepsilon$ where $\mathbb{Q}_+$ is the set of positive rational numbers. Then $\mathbb{P}(S) = 1$ and for each rational $\varepsilon > 0$ we have

$$\limsup_{n\to\infty}\sup_{\theta\in\Theta}|Q_n(\theta;\omega)-Q(\theta)| \leq \varepsilon \tag{24}$$

for all $\omega \in S$. As $\varepsilon \in \mathbb{Q}_+$ is arbitrary, we have shown that:

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}|Q_n(\theta;\omega)-Q(\theta)| = 0 \tag{25}$$

for all $\omega \in S$. ■

How do we show a collection of functions has finite bracketing numbers? The following result uses compactness of Θ, continuity, and dominance assumptions.

Lemma 3. *Let the following hold:*
(i) Θ *is compact*
(ii) $m(X_t;\theta)$ *is continuous in* θ *for all* X_t
(iii) $\mathrm{E}[\sup_{\theta\in\Theta}|m(X_t;\theta)|] < \infty$.
Then: $\mathcal{M} = \{m(X_t,\theta) : \theta \in \Theta\}$ *has finite bracketing numbers. If, in addition,*
(iv) $X_1,\ldots,X_n$ *are IID or SSE, then:* $\sup_{\theta\in\Theta}|\frac{1}{n}\sum_{t=1}^n m(X_t,\theta) - \mathrm{E}[m(X_t,\theta)| \to_{a.s.} 0$.

Proof. Fix any $\delta > 0$. As Θ is compact, we can cover Θ with finitely many open balls of radius δ centered at $\theta_1,\ldots,\theta_J$. For each $j = 1,\ldots,J$, define

$$l_{\delta,j}(\cdot) = \inf_{\theta\in\Theta:\|\theta-\theta_j\|\leq\delta} m(\cdot;\theta) \quad \text{and} \quad u_{\delta,j}(\cdot) = \sup_{\theta\in\Theta:\|\theta-\theta_j\|\leq\delta} m(\cdot;\theta)\,, \tag{26}$$

so that $l_{\delta,j}(\cdot) \leq m(\cdot;\theta) \leq u_{\delta,j}(\cdot)$ holds for each θ with $\|\theta-\theta_j\| \leq \delta$. Note that the inf and sup are always finite by (i) and (ii).

Let $\varepsilon(\delta) = \max_{1\leq j\leq J} \mathrm{E}[u_{\delta,j}(X_t) - l_{\delta,j}(X_t)]$. We have shown that $N_{[\,]}(\varepsilon(\delta), \mathcal{M}) \leq J < \infty$. It remains to show that $\varepsilon(\delta) \to 0$ as $\delta \to 0$. This will ensure that for any $\epsilon > 0$ we can choose a δ such that $\varepsilon(\delta) \leq \epsilon$, and hence that $N_{[\,]}(\epsilon, \mathcal{M}) < \infty$ for each $\epsilon > 0$.

Let $M_\delta(\cdot) = \max_{1\leq j\leq J}(u_{\delta,j}(\cdot) - l_{\delta,j}(\cdot))$. We may use (i) and (ii) to deduce that $M_\delta(X_t) \to 0$ as $\delta \to 0$ for each X_t (Exercise: use the fact that a continuous function on a compact set is uniformly continuous to show this formally). Also notice that $|M_\delta(X_t)| \leq 2\sup_{\theta\in\Theta}|m(X_t;\theta)|$. Then by (iii) we may apply the dominated convergence theorem to obtain:

$$\lim_{\delta\to 0}\varepsilon(\delta) \leq \lim_{\delta\to 0}\mathrm{E}[M_\delta(X_t)] = \mathrm{E}[\lim_{\delta\to 0} M_\delta(X_t)] = 0 \tag{27}$$

as required. ∎

Combining Lemmas 1, 2 and 3 gives the following consistency result.

Theorem 2 (Consistency of M-estimators)**.** *Let the following hold:*
(i) $X_1, \ldots, X_n$ *are IID or SSE*
(ii) Θ *is compact*
(iii) $m(X_t;\theta)$ *is continuous in* θ *for all* X_t
(iv) $\mathrm{E}[\sup_{\theta\in\Theta}|m(X_t;\theta)|] < \infty$
(v) $Q(\theta_0) > Q(\theta)$ *for all* $\theta \in \Theta$ *with* $\theta \neq \theta_0$*.*
Then: $\hat{\theta} \to_p \theta_0$ *as* $n \to \infty$*.*

Proof. By Theorem 1 we just need to verify "clean maximum" and "uniform convergence".

We use Lemma 1 to verify "clean maximum". By conditions (ii) and (v), it is enough to show that Q is continuous under the stated conditions. To verify continuity of Q, take any $\theta^* \in \Theta$ and let $(\theta_n)_{n\in\mathbb{N}} \subset \Theta$ be a sequence such that $\|\theta_n - \theta^*\| \to 0$ as $n \to \infty$. By condition (iii) we know that $\lim_{n\to\infty} m(X_t;\theta_n) = m(X_t;\theta^*)$ for all X_t. Then by condition (iv) we may apply the dominated convergence theorem to deduce

$$\lim_{n\to\infty} Q(\theta_n) = \lim_{n\to\infty}\mathrm{E}[m(X_t;\theta_n)] = \mathrm{E}[\lim_{n\to\infty} m(X_t;\theta_n)] = \mathrm{E}[m(X_t;\theta^*)] = Q(\theta^*)\,, \tag{28}$$

which verifies continuity of Q. Therefore "clean maximum" holds.

Conditions (ii)–(iv) give finite bracketing numbers by Lemma 3. Moreover, $\mathcal{M} \subset L^1$ by (iv). This, together with (i), gives "uniform convergence" by Lemma 2. ∎

3.2 Consistency of GMM Estimators

We're going to apply Theorem 1 to establish consistency of the GMM estimator. This requires verifying "clean maximum" and "uniform convergence".

To apply Lemma 2, we need some notation. Write

$$g(X_t;\theta) = \begin{pmatrix} g_1(X_t;\theta) \\ g_2(X_t;\theta) \\ \vdots \\ g_K(X_t;\theta) \end{pmatrix}. \tag{29}$$

Then with this notation,

$$g_n(\theta) - g(\theta) = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^n g_1(X_t;\theta) - \mathrm{E}[g_1(X_t;\theta)] \\ \frac{1}{n}\sum_{t=1}^n g_2(X_t;\theta) - \mathrm{E}[g_2(X_t;\theta)] \\ \vdots \\ \frac{1}{n}\sum_{t=1}^n g_K(X_t;\theta) - \mathrm{E}[g_K(X_t;\theta)] \end{pmatrix}. \tag{30}$$

We will use Lemma 2 to ensure that each entry of $g_n(\theta) - g(\theta)$ converges in probability to zero (uniformly in θ), and hence $\sup_{\theta\in\Theta} \|g_n(\theta) - g(\theta)\| \to_p 0$. Let $\mathcal{G} = \{g_k(\cdot;\theta) : \theta \in \Theta, 1 \le k \le K\}$.

Theorem 3 (Consistency of GMM estimators)**.** *Let the following hold:*
(i) $X_1, \ldots, X_n$ *are IID or SSE*
(ii) Θ *is compact*
(iii) $g(\theta)$ *is continuous*
(iv) $\widehat{W} \to_p W$ *where* W *is positive definite and symmetric*
(v) $g(\theta) = 0$ *if and only if* $\theta = \theta_0$
(vi) $\mathcal{G}$ *has finite bracketing numbers.*
Then: $\hat{\theta} \to_p \theta_0$ *as* $n \to \infty$*.*

Proof. We verify the conditions of Theorem 1.

Continuity of $Q(\theta)$ follows from continuity of $g(\theta)$ and positive-definiteness of W. Therefore "clean maximum" holds by Lemma 1 (under conditions (ii)–(v)).

We now verify "uniform convergence". Step 1: we show $\sup_{\theta\in\Theta} \|g_n(\theta) - g(\theta)\| \to_p 0$. Assumption (vi) implies that each of the K component functions in $g(X_t;\theta)$ has finite bracketing numbers. We may then apply Lemma 2 to deduce that each entry of $g_n(\theta) - g(\theta)$ converges in probability to zero (uniformly in θ), and hence

$$\sup_{\theta\in\Theta} \|g_n(\theta) - g(\theta)\| \to_p 0. \tag{31}$$

Before proceeding, we note that as g is continuous and Θ is compact, we also have:

$$\sup_{\theta\in\Theta}\|g(\theta)\| < \infty\,. \tag{32}$$

Combining (31) and (32) gives:

$$\sup_{\theta\in\Theta}\|g_n(\theta)\| \le \sup_{\theta\in\Theta}\|g_n(\theta)-g(\theta)\| + \sup_{\theta\in\Theta}\|g(\theta)\| = o_p(1) + \sup_{\theta\in\Theta}\|g(\theta)\| = O_p(1)\,. \tag{33}$$

Step 2: we show $\sup_{\theta\in\Theta}|Q_n(\theta)-Q(\theta)| \to_p 0$. Adding and subtracting terms:

$$\begin{aligned}
2Q(\theta)-2Q_n(\theta) &= g_n(\theta)'\widehat{W}g_n(\theta) - g(\theta)'Wg(\theta) && (34)\\
&= g_n(\theta)'(\widehat{W}-W)g_n(\theta) + g_n(\theta)'Wg_n(\theta) - g(\theta)'Wg(\theta) && (35)\\
&= g_n(\theta)'(\widehat{W}-W)g_n(\theta) + \big(g_n(\theta)-g(\theta)\big)'W\big(g_n(\theta)+g(\theta)\big)\,. && (36)
\end{aligned}$$

Notice that for any K-vectors x, y and $K\times K$ matrix A we have

$$|x'Ay| \le \|x\|\|y\|\|A\| \tag{37}$$

where $\|x\|$ and $\|y\|$ are the Euclidean norms of x and y and $\|A\|$ is the spectral norm (largest singular value) of A. Applying the triangle inequality then inequality (37) to (36) yields:

$$\begin{aligned}
\sup_{\theta\in\Theta} 2|Q_n(\theta)-Q(\theta)| &\le \sup_{\theta\in\Theta}\big|g_n(\theta)'(\widehat{W}-W)g_n(\theta)\big| \\
&\qquad + \sup_{\theta\in\Theta}\big|\big(g_n(\theta)-g(\theta)\big)'W\big(g_n(\theta)+g(\theta)\big)\big| && (38)\\
&\le \underbrace{\Big(\sup_{\theta\in\Theta}\|g_n(\theta)\|\Big)^2}_{=O_p(1)\text{ by }(33)} \times \underbrace{\|\widehat{W}-W\|}_{=o_p(1)\text{ by (iv)}} \\
&\qquad + \underbrace{\sup_{\theta\in\Theta}\big\|\big(g_n(\theta)-g(\theta)\big)\big\|}_{=o_p(1)\text{ by }(31)} \times \underbrace{\sup_{\theta\in\Theta}\|g_n(\theta)+g(\theta)\|}_{=O_p(1)\text{ by }(32)\text{ and }(33)} \times \|W\| && (39)\\
&= O_p(1)\times o_p(1) + o_p(1)\times O_p(1)\times \text{constant} = o_p(1)\,, && (40)
\end{aligned}$$

which verifies "uniform convergence". ∎

3.3 Consistency of SMM Estimators

Consistency for SMM requires special treatment because of the additional noise introduced by the simulation draws. Let's suppose that the simulated data $X_1^\theta, \ldots, X_m^\theta$ are generated as functions of i.i.d. draws $\varepsilon_1, \ldots, \varepsilon_m$ which represent the "shocks" used to simulate the data. That is,

$$X_s^\theta = a(\varepsilon_s, \theta) \tag{41}$$

for each $1 \leq s \leq m$ and each $\theta \in \Theta$. We expand the probability space to jointly accommodate the true data $X_1, \ldots, X_n$ and the simulated draws $\varepsilon_1, \ldots, \varepsilon_m$.[1] All probability statements we make in reference to SMM are to be understood with respect to the joint probability law of the data and simulated draws. As the sample size n gets large, we will be taking $m \to \infty$ also. If we don't, the simulation error will eventually dominate and the SMM estimator will not converge.

We again establish consistency by verifying "clean maximum" and "uniform convergence".

Theorem 4 (Consistency of SMM estimators)**.** *Let the following hold:*
(i) Θ *is compact*
(ii) $\gamma(\theta)$ *is continuous in* θ
(iii) $\sup_{\theta \in \Theta} \|\gamma_m(\theta) - \gamma(\theta)\| = o_p(1)$
(iv) $g_n \to_p g_0$ *and* $\widehat{W} \to_p W$ *where* W *is positive definite and symmetric*
(v) $\gamma(\theta) = g_0$ *if and only if* $\theta = \theta_0$.
Then: $\hat{\theta} \to_p \theta_0$ *as* $n \to \infty$.

Note that in (iii) we explicitly assume the simulated moments converge (uniformly) to the moment function $\gamma(\theta)$ as the number of simulations increases. This can be verified under more primitive conditions by applying Lemma 2, substituting $\varepsilon_1, \ldots, \varepsilon_m$ for $X_1, \ldots, X_n$ and $\gamma(a(\varepsilon_s, \theta); \theta)$ for $m(X_t, \theta)$.

Proof. By Theorem 1 we just need to verify "clean maximum" and "uniform convergence".

We use Lemma 1 to verify "clean maximum", noting Θ is compact (by (i)), $Q(\theta)$ is continuous (by (ii) and finiteness of W), and $Q(\theta_0) > Q(\theta)$ for any $\theta \neq \theta_0$ (by (v) and positive-definiteness of W).

We verify "uniform convergence" by similar arguments to the proof of Lemma 3. Adding and subtracting terms:

$$\begin{aligned} 2Q(\theta) - 2Q_n(\theta) &= (g_n - \gamma_m(\theta))'\widehat{W}(g_n - \gamma_m(\theta)) - (g_0 - \gamma(\theta))'W(g_0 - \gamma(\theta)) && (42) \\ &= (g_n - \gamma_m(\theta))'(\widehat{W} - W)(g_n - \gamma_m(\theta)) \\ &\quad + \big(g_n - g_0 + \gamma(\theta) - \gamma_m(\theta)\big)'W\big(g_n - \gamma_m(\theta) + g_0 - \gamma(\theta)\big)\,. && (43) \end{aligned}$$

Conditions (iii) and (iv) imply that

$$\sup_{\theta \in \Theta} \|g_n - g_0 + \gamma(\theta) - \gamma_m(\theta)\| \leq \|g_n - g_0\| + \sup_{\theta \in \Theta} \|\gamma_m(\theta) - \gamma(\theta)\| = o_p(1) \tag{44}$$

[1] This is achieved by joining the σ-fields of the two and using the fact that the simulation draws are totally independent of the data.

and, moreover,

$$\sup_{\theta\in\Theta}\|g_n-\gamma_m(\theta)\| \leq \sup_{\theta\in\Theta}\|g_0-\gamma(\theta)\| + \sup_{\theta\in\Theta}\|g_n-g_0+\gamma(\theta)-\gamma_m(\theta)\| = O_p(1) \tag{45}$$

because $\sup_{\theta\in\Theta}\|g_0-\gamma(\theta)\| < \infty$ by conditions (i) and (ii). Applying the triangle inequality then inequality (37) to (43), we obtain

$$\begin{aligned}
\sup_{\theta\in\Theta} 2|Q_n(\theta)-Q(\theta)| \quad &\leq \quad \underbrace{\left(\sup_{\theta\in\Theta}\|g_n-\gamma_m(\theta)\|\right)^2}_{=O_p(1) \text{ by } (45)} \times \underbrace{\|\widehat{W}-W\|}_{=o_p(1) \text{ by (iv)}} \\
&\qquad + \underbrace{\sup_{\theta\in\Theta}\|g_n-g_0+\gamma(\theta)-\gamma_m(\theta)\|}_{=o_p(1) \text{ by } (44)} \\
&\qquad \times \underbrace{\sup_{\theta\in\Theta}\|g_n-\gamma_m(\theta)+g_0-\gamma(\theta)\|}_{=O_p(1) \text{ by } (45)} \times \|W\| \qquad (46) \\
&= \quad O_p(1)\times o_p(1) + o_p(1)\times O_p(1)\times \text{constant} = o_p(1)\,, \qquad (47)
\end{aligned}$$

which verifies "uniform convergence". ∎

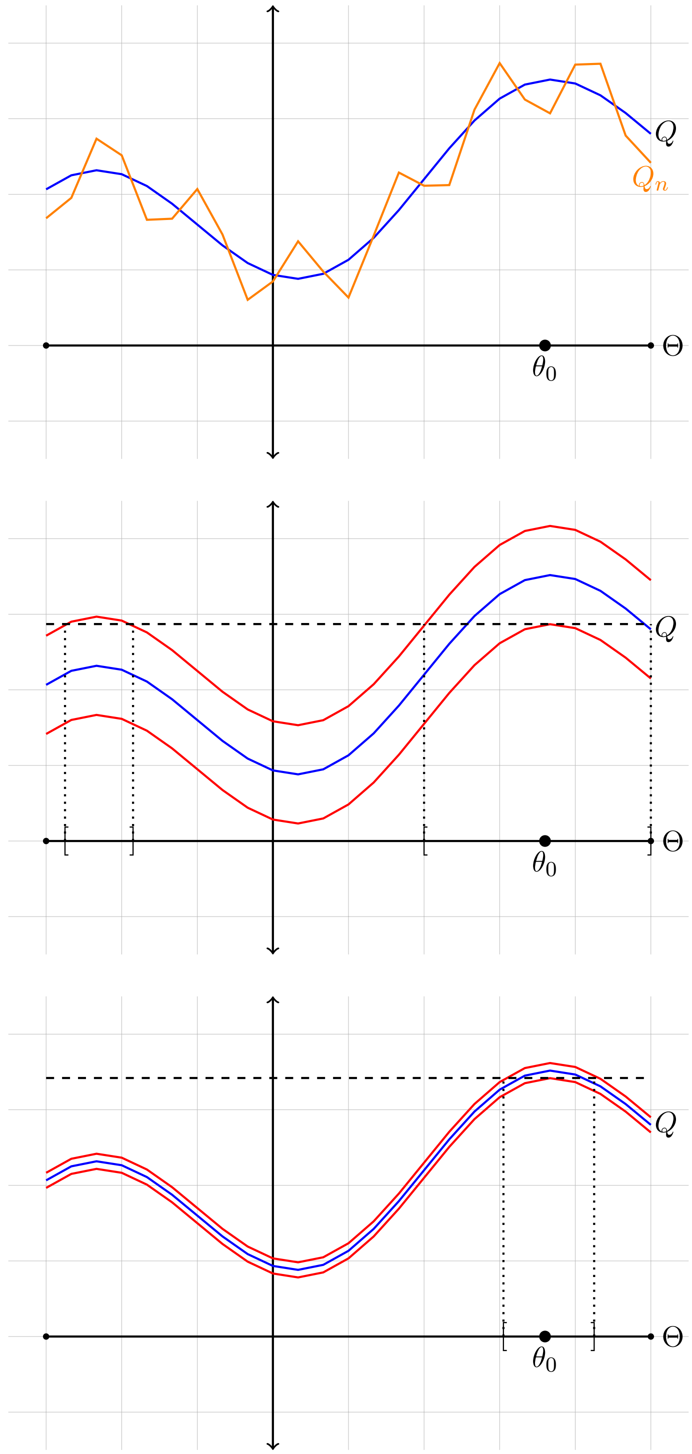

Figure 1: Consistency of Extremum Estimators. When $Q_n(\theta)$ lies uniformly in $[Q(\theta) - \epsilon, Q(\theta) + \epsilon]$ we know that $\hat{\theta}$ must be in the set $\{\theta : Q(\theta) \geq Q(\theta_0) - \epsilon\}$. Provided clean maximum holds, this set becomes a shrinking interval around θ_0 as ϵ decreases.

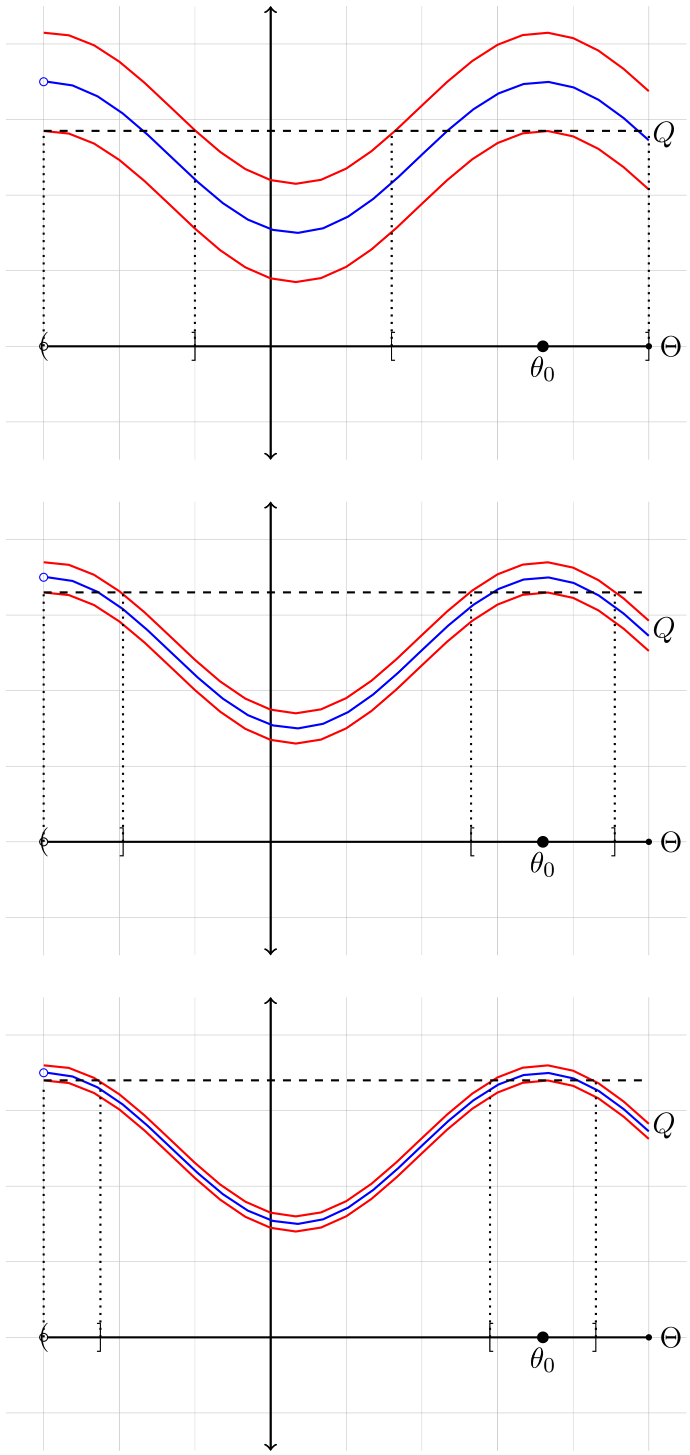

Figure 2: Necessity of Clean Maximum.