Note for Asymptotic Analysis in Econometrics

Part I: Consistency

Fu Ouyang

December 21, 2016

1 Introduction

Definition 1.1. An estimator $\widehat{\theta}_n$ is called an *extremum estimator* if there is an objective function $Q_n(\theta)$ such that

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} Q_n(\theta)$$

where Θ is the parameter space.

This note is designed to collect and introduce main results on establishing the consistency of extremum estimators, as well as a guide to recent developments. All results are assumed to be developed for i.i.d. data unless otherwise remarked.

2 Consistency

The basic idea of showing consistency is that if $Q_n(\theta)$ converges in probability to $Q_0(\theta)$ for every $\theta \in \Theta$, and $Q_0(\theta)$ is (uniquely) maximized at the true parameter $\theta_0 \in \Theta$, then the limit of the maximum $\widehat{\theta}_n$ should be the maximum θ_0 of the limit, under some conditions for interchanging the maximization and limiting operators. Note that the condition that $Q_0(\theta)$ has a unique maximum at the true parameter θ_0 is related to identification which is generally an essential part of a consistency theorem.

Definition 2.1. $Q_n(\theta)$ *converges uniformly in probability* to $Q_0(\theta)$ means

$$\sup_{\theta\in\Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$$

Theorem 2.1 (fundamental consistency theorem, Theorem 2.1 of Newey and McFadden (1994))**.** *If there is a function $Q_0(\theta)$ such that (i) $Q_0(\theta)$ is uniquely maximized at $\theta_0 \in \Theta$; (ii) Θ is compact; (iii) $Q_0(\theta)$ is continuous; (iv) $Q_n(\theta)$ converges uniformly in probability to $Q_0(\theta)$, then $\widehat{\theta}_n \xrightarrow{p} \theta_0$.*[1]

Proof. For any $\varepsilon > 0$, we have with probability approaching one (w.p.a.1) (a) $Q_n(\widehat{\theta}_n) > Q_n(\theta_0) - \varepsilon/3$ by definition of $\widehat{\theta}_n$; (b) $Q_0(\widehat{\theta}_n) > Q_n(\widehat{\theta}_n) - \varepsilon/3$ by (iv); (c) $Q_n(\theta_0) > Q_0(\theta_0) - \varepsilon/3$ by (iv). Therefore, w.p.a.1 we have

$$Q_0(\widehat{\theta}_n) > Q_n(\widehat{\theta}_n) - \varepsilon/3 > Q_n(\theta_0) - 2\varepsilon/3 > Q_0(\theta_0) - \varepsilon$$

Thus, for any $\varepsilon > 0$, $Q_0(\widehat{\theta}_n) > Q_0(\theta_0) - \varepsilon$ w.p.a.1. Let $\mathcal{N}$ be any open subset of Θ such that $\theta_0 \in \mathcal{N}$. Note that (ii) implies $\Theta \cap \mathcal{N}^c$ is compact. Then, by (i) and (iii), we have $\sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_0(\theta) = Q_0(\theta') < Q_0(\theta_0)$ for some $\theta' \in \Theta \cap \mathcal{N}^c$. Let $\varepsilon = Q_0(\theta_0) - \sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_0(\theta)$. It follows that w.p.a.1

$$Q_0(\widehat{\theta}_n) > Q_0(\theta_0) - \varepsilon = \sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_0(\theta)$$

and hence $\widehat{\theta}_n \in \mathcal{N}$. □

Remark. The conditions of this theorem can be slightly weaker:

1. $\widehat{\theta}_n$ may not be the maximum of the objective function, instead we just need $Q_n(\widehat{\theta}_n) \geq \sup_{\theta \in \Theta} Q_n(\theta) + o_p(1)$.
2. (iii) can be changed to upper-semicontinuity of $Q_0(\theta)$ since $\Theta \cap \mathcal{N}^c$ is compact and a upper-semicontinuous $Q_0(\theta)$ attains its supremum in $\Theta \cap \mathcal{N}^c$.
3. (iv) can be relaxed to $Q_n(\theta_0) \xrightarrow{p} Q_0(\theta_0)$ and for all $\varepsilon > 0$, $Q_n(\theta) < Q_0(\theta) + \varepsilon$ for all $\theta \in \Theta$ w.p.a.1.

Remark. This result is quite general, applying to any topological space. Hence, Θ can be infinite-dimensional.

Remark. Condition (i) is called identification condition that is related to identification in the usual sense. Identification means that the distribution of the data at the true parameter is different from that at any other possible parameter value. Identification is necessary for identification condition, but not in general sufficient.

Remark. If (iv) is replaced by $\sup_{\theta \in \Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{a.s.} 0$, we can show $\widehat{\theta}_n \xrightarrow{a.s.} \theta_0$ with similar argument.

[1] To avoid measurability complications it will be assumed that probability statements are for outer probability.

Remark. Condition (i) and (ii) are called "substantive" condition that should be verified case-by-case. Condition (iii) and (iv) are "regularity" conditions and typically cited as "the standard regularity conditions". (iii) is a weak condition, to verify it, some moment conditions may be needed for using LLN. (iv) can be verified by some uniform law of large numbers or using the property "stochastic equicontinuity".

For many estimators maximizing sample averages, the uniform convergence condition (iv) can be verified by the following *uniform law of large numbers*. Let $a(z,\theta)$ be a matrix of functions of an observation z and the parameter θ, and for a matrix $A = [a_{jk}]$, let $\|A\| = (\sum_{j,k} a_{jk}^2)^{1/2}$ be the Euclidean norm.

Lemma 2.2 (Lemma 2.4 of Newey and McFadden (1994)). *If the data are i.i.d.,* Θ *is compact,* $a(z,\theta)$ *is continuous at each* $\theta \in \Theta$ *with probability* 1*, and there is* $d(z)$ *with* $\|a(z,\theta)\| \leq d(z)$ *for all* $\theta \in \Theta$ *and* $\mathbb{E}[d(z)] < \infty$*, then* $\mathbb{E}[a(z,\theta)]$ *is continuous and* $\sup_{\theta\in\Theta} \|n^{-1}\sum_{i=1}^n a(z_i,\theta) - \mathbb{E}[a(z,\theta)]\| \xrightarrow{p} 0$.[2]

The conditions of this lemma are weak. First, it allows for $a(z,\theta)$ to not be continuous on all of Θ given z. Also, the conclusion remains true if the i.i.d. hypothesis is relaxed to strict stationarity and ergodicity of z_i. To verify the "dominance condition", where $d(z)$ is called dominating function, one can use tools such as the mean-value theorem, Cauchy-Schwartz inequality, etc.

Definition 2.2 (Stochastic equicontinuity). For every $\varepsilon, \eta > 0$ there exists a sequence of random variables $\Delta_n(\varepsilon,\eta)$ and a sample size $n_0(\varepsilon,\eta)$ such that for $n > n_0(\varepsilon,\eta)$, $\Pr(|\Delta_n(\varepsilon,\eta)| > \varepsilon) < \eta$ and for each θ there is an open set $\mathcal{N}(\theta,\varepsilon,\eta)$ containing θ with

$$\sup_{\theta'\in\mathcal{N}(\theta,\varepsilon,\eta)} |Q_n(\theta') - Q_n(\theta)| \leq \Delta_n(\varepsilon,\eta),\ \forall n > n_0(\varepsilon,\eta)$$

Remark. $\Delta_n = o_p(1)$ acts like a "random epsilon", bounding the effect of changing θ on $Q_n(\theta)$.

Lemma 2.3 (Lemma 2.8 of Newey and McFadden (1994)). *Suppose* Θ *is compact and* $Q_0(\theta)$ *is continuous. Then* $\sup_{\theta\in\Theta}|Q_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$ *if and only if* $Q_n(\theta) \xrightarrow{p} Q_0(\theta)$ *for all* $\theta \in \Theta$ *and* $Q_n(\theta)$ *is stochastically equicontinuous.*

Proof. By Lemma A.1 it suffices to show the result with $R_n(\theta) = Q_n(\theta) - Q_0(\theta)$ replacing $Q_n(\theta)$. Consider $\varepsilon, \eta > 0$ with $\mathcal{N}(\theta,\varepsilon/2,\eta)$, $\Delta_n(\varepsilon/2,\eta)$ and $n_0(\varepsilon/2,\eta)$. By Θ compact there exists a finite open covering $\{\mathcal{N}(\theta_j)\}_{j=1}^J$ of $\mathcal{N}(\theta,\varepsilon/2,\eta)$, then by triangle inequality,

$$\sup_{\theta\in\Theta}|R_n(\theta)| \leq \sup_j |R_n(\theta_j)| + \sup_{\theta\in\Theta}\sup_{j,\theta\in\mathcal{N}(\theta_j)} |R_n(\theta) - R_n(\theta_j)| \leq o_p(1) + \Delta_n(\varepsilon/2,\eta)$$

[2]This lemma is implied by Lemma 1 of Tauchen (1985), and it has similar assumptions of Wald's (1949) consistency proof.

Then for n sufficiently large, we have

$$\Pr(\sup_{\theta\in\Theta} |R_n(\theta)| > \varepsilon) \leq \Pr(o_p(1) > \varepsilon/2) + \Pr(\Delta_n(\varepsilon/2, \eta) > \varepsilon/2) < \eta$$

Conversely, first note that uniform convergence in probability trivially implies convergence in probability. Also, let $\Delta_n(\varepsilon, \eta) = 2\sup_{\theta\in\Theta} |R_n(\theta)| = o_p(1)$. We have for any θ and its open neighborhood $\mathcal{N}$, by triangle inequality,

$$\sup_{\theta'\in\mathcal{N}} |R_n(\theta') - R_n(\theta)| \leq \Delta_n(\varepsilon, \eta)$$

□

Remark. This results can be applied to the case in which $Q_0(\theta)$ is replaced by some nonrandom function of θ and n, $\overline{Q}_n(\theta)$. In this case, $\overline{Q}_n(\theta)$ should be assumed to be equicontinuous. See Theorem 2.1 of Newey (1991).

One useful sufficient condition for uniform convergence that is motivated by the stochastic equicontinuity property is a global "Lipschitz"-type condition.

Lemma 2.4 (Lemma 2.9 of Newey and McFadden (1994))**.** *If Θ is compact, $Q_0(\theta)$ is continuous, $Q_n(\theta) \xrightarrow{p} Q_0(\theta)$ for all $\theta \in \Theta$, and there is $\alpha > 0$ and $B_n = O_p(1)$ such that for all $\theta', \theta \in \Theta$, $|Q_n(\theta') - Q_n(\theta)| \leq B_n \|\theta' - \theta\|^\alpha$, then $\sup_{\theta\in\Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$.*

Proof. By Lemma 2.3 it suffices to show stochastic equicontinuity. Pick $\varepsilon, \eta > 0$. By $B_n = O_p(1)$ there exists $M(n_0) > 0$ such that $\Pr(|B_n| > M(n_0)) < \eta$ for all $n > n_0$. Let $\Delta_n = B_n\varepsilon/M(n_0)$ and $\mathcal{N} = \{\theta' : \|\theta' - \theta\|^\alpha \leq \varepsilon/M(n_0)\}$. Then $\Pr(|\Delta_n| > \varepsilon) = \Pr(|B_n| > M(n_0)) < \eta$ for all $n > n_0$. Hence, $\Delta_n = o_p(1)$ and $\sup_{\theta'\in\mathcal{N}} |Q_n(\theta') - Q_n(\theta)| \leq B_n \|\theta' - \theta\|^\alpha \leq \Delta_n$. □

Remark. A more general Lipschitz condition is $|Q_n(\theta') - Q_n(\theta)| \leq B_n h(\|\theta' - \theta\|)$, where $h : [0, \infty) \mapsto [0, \infty)$ is a function continuous at 0 with $h(0) = 0$. In fact, by Lemma 2.3, it suffices to show the equicontinuity of $Q_n(\theta)$. Consider $\varepsilon, \eta > 0$. Since $B_n = O_p(1)$, there is $M > 0$ such that for all n, $\Pr(B_n > \varepsilon M) < \eta$, so one can let $\Delta_n(\varepsilon, \eta) = B_n/M$. Choose $\delta > 0$ sufficiently small that $h(d) < 1/M$ for all $0 \leq d < \delta$ and let $\mathcal{N}(\theta, \varepsilon, \eta) = \{\tilde{\theta} \in \Theta : \|\tilde{\theta} - \theta\| < \delta\}$. Then $\sup_{\mathcal{N}} |Q_n(\tilde{\theta}) - Q_n(\theta)| \leq B_n \sup_{0\leq d<\delta} h(d) \leq \Delta_n(\varepsilon, \eta)$. See Corollary 2.2 of Newey (1991).

When the objective function is concave, it is possible to avoid assuming compactness of Θ. Intuitively, concavity prevents the objective function from "turning up" as the parameter moves far away from the truth.

Theorem 2.5 (Theorem 2.7 of Newey and McFadden (1994))**.** *If there is a function $Q_0(\theta)$ such that (i) $Q_0(\theta)$ is uniquely maximized θ_0; (ii) $\theta_0 \in int(\Theta)$, Θ is a convex set, and $\widehat{Q}_n(\theta)$ is concave; and (iii) $\widehat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta)$ for all $\theta \in \Theta$, then exists w.p.a.1 and $\widehat{\theta}_n \xrightarrow{p} \theta$.*[3]

[3] A concave function is continuous on the interior of its (convex) domain.

In addition to allowing for noncompact Θ, Theorem 2.5 only requires pointwise convergence, which is possible because pointwise convergence of concave functions implies uniform convergence[4]. Note that this theorem contains the additional conclusion that $\widehat{\theta}$ exists w.p.a.1, which is needed because of noncompactness of Θ.

The following lemma slightly generalizes Lemma 2.4.

Lemma 2.6 (Lemma A.2 of Newey and Powell (2003)). *if (i) Θ is a compact subset of a space with norm $\|\cdot\|$; (ii) $Q_n(\theta) \xrightarrow{p} Q_0(\theta)$ for all $\theta \in \Theta$; (iii) there is a $B_n = O_p(1)$ and $\alpha > 0$ such that for all $\theta', \theta \in \Theta$, $|Q_n(\theta') - Q_n(\theta)| \leq B_n \|\theta' - \theta\|^\alpha$, then $Q_0(\theta)$ is continuous and* $\sup_{\theta\in\Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$.

Proof. Uniform convergence in probability is guaranteed by Lemma 2.3. To show continuity of $Q_0(\theta)$ consider any fixed $\theta \in \Theta$ and $\varepsilon > 0$. By $B_n = O_p(1)$ there exists $M > 0$ such that $\Pr(B_n \leq M) > 0$ for all n. Let $\Delta = (\varepsilon/2M)^{1/\alpha}$. Note that for all θ' with $\|\theta' - \theta\| \leq \Delta$ we have $|Q_n(\theta') - Q_n(\theta)| \leq B_n\Delta^\alpha = \varepsilon B_n/2M \leq \varepsilon/2$ with positive probability. By triangle inequality and (ii)

$$|Q_0(\theta') - Q_0(\theta)| \leq |Q_n(\theta') - Q_n(\theta)| + |Q_0(\theta') - Q_n(\theta')| + |Q_0(\theta) - Q_n(\theta)| \leq \varepsilon$$

with positive probability with n sufficiently large. Note that $|Q_0(\theta') - Q_0(\theta)|$ is constant and θ is arbitrary. Hence, $Q_0(\theta)$ is continuous. $\square$

The following theorem improved Theorem 2.1 of Newey and McFadden (1994).

Theorem 2.7 (Lemma A.1 of Newey and Powell (2003)). *Suppose (i) $Q_0(\theta)$ is uniquely minimized at $\theta_0 \in \Theta$; (ii) Θ is compact, $Q_0(\theta)$ and $Q_n(\theta)$ are both continuous[5], and* $\sup_{\theta\in\Theta} |Q_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0$*; (iii) Θ_n are compact subsets of Θ such that for any $\theta \in \Theta$ there exists $\widetilde{\theta}_n \in \Theta_n$ such that $\widetilde{\theta}_n \xrightarrow{p} \theta$. then $\widehat{\theta}_n \xrightarrow{p} \theta_0$.*

Proof. Consider any open set $\mathcal{N}$ containing θ_0. By compactness of Θ, continuity of $Q_0(\theta)$, and $Q_0(\theta)$ having a unique minimum at θ_0, define

$$\Delta = \inf_{\theta\in\Theta\cap\mathcal{N}^c} Q_0(\theta) - Q_0(\theta_0) > 0$$

By (iii) there is $\widetilde{\theta}_n \in \Theta_n$ such that $\widetilde{\theta}_n \xrightarrow{p} \theta_0$. By the uniform convergence condition in (ii), we have $Q_0(\widehat{\theta}_n) - Q_n(\widehat{\theta}_n) \leq \Delta/4$ and $Q_n(\widetilde{\theta}_n) - Q_0(\widetilde{\theta}_n) \leq \Delta/4$ w.p.a.1. Since by definition $Q_n(\widehat{\theta}_n) \leq Q_n(\widetilde{\theta}_n)$, $Q_0(\widehat{\theta}_n) \leq Q_0(\widetilde{\theta}_n)+\Delta/2$ w.p.a.1. Furthermore, by the definition of $\widetilde{\theta}_n$ and continuity of $Q_0(\theta)$, $Q_0(\widetilde{\theta}_n) \leq Q_0(\theta_0)+\Delta/2$ w.p.a.1. Then by summing the two inequalities and substracting $Q_0(\widetilde{\theta}_n)$ from both sides, $Q_0(\widehat{\theta}_n) \leq Q_0(\theta_0) + \Delta = \inf_{\theta\in\Theta\cap\mathcal{N}^c} Q_0(\theta)$ w.p.a.1 and so $\widehat{\theta}_n \in \mathcal{N}$ w.p.a.1. Therefore, $\widehat{\theta}_n \xrightarrow{p} \theta_0$ follows from that $\mathcal{N}$ is any open set containing θ_0. $\square$

[4]See the proof in Newey and McFadden (1994) for details.

[5]Why do we need $Q_n(\theta)$ to be continuous?

Remark. Θ_n is often an approximating set that can be thought of as a finite-dimensional subset of Θ.

Remark. Theorem 2.7 is a convergence in probability version of Theorem 0 of Gallant and Nychka (1987). Lemma 2.6 is a slightly improved version of Corollary 2.2 of Newey (1991).

Theorem 2.8 (Theorem 0 of Gallant and Nychka (1987))**.**

Proof. □

References

Gallant, A. R. and Douglas W. Nychka (1987). "Semi-nonparametric maximum likelihood estimation." In: *Econometrica* 55.2, pp. 363–390.

Newey, Whitney K. (1991). "Uniform Convergence in Probability and stochsatic equicontinuity." In: *Econometrica* 59.4, pp. 1161–1167.

Newey, Whitney K and Daniel McFadden (1994). "Large sample estimation and hypothesis testing." In: *Handbook of econometrics* 4, pp. 2111–2245.

Newey, Whitney K. and James L. Powell (2003). "Instrumental variable estimation of nonparametric models." In: *Econometrica* 71.5, pp. 1565–1578.

A Auxilliary Lemmas

The following Lemma is a little more general than required.

Lemma A.1 (Lemma A.1 of Newey (1991))**.** *Define $R_n(\theta) = Q_n(\theta) - \overline{Q}_n(\theta)$. If $\overline{Q}_n(\theta)$ is equicontinuous and Θ is compact, then $Q_n(\theta)$ is stochastic equicontinuous if and only if $R_n(\theta)$ is stochastic equicontinuous.*

Proof. Note that for any $\theta, \theta' \in \Theta$ we have, by triangle inequality,

$$|R_n(\theta') - R_n(\theta)| \leq |Q_n(\theta') - Q_n(\theta)| + \left|\overline{Q}_n(\theta') - \overline{Q}_n(\theta)\right| \tag{A.1}$$

and

$$|Q_n(\theta') - Q_n(\theta)| \leq |R_n(\theta') - R_n(\theta)| + \left|\overline{Q}_n(\theta') - \overline{Q}_n(\theta)\right| \tag{A.2}$$

If $Q_n(\theta)$ is stochastic equicontinuous, there exists $\mathcal{N}(\theta, \varepsilon/2, \eta)$, $\Delta_n(\varepsilon/2, \eta)$ and $n_0(\varepsilon/2, \eta)$ such that

$$\sup_{\theta' \in \mathcal{N}(\theta, \varepsilon/2, \eta)} |Q_n(\theta') - Q_n(\theta)| \leq \Delta_n(\varepsilon/2, \eta),\ \forall n > n_0(\varepsilon/2, \eta)$$

Let $\mathcal{N}'(\theta, \varepsilon/2, \eta) \subset \mathcal{N}(\theta, \varepsilon/2, \eta)$ be an open set such that

$$\sup_{n, \theta' \in \mathcal{N}'(\theta, \varepsilon/2, \eta)} \left| \overline{Q}_n(\theta') - \overline{Q}_n(\theta) \right| \leq \varepsilon/2$$

Then by (A.1) for $n > n_0(\varepsilon/2, \eta)$

$$\sup_{\theta' \in \mathcal{N}'(\theta, \varepsilon/2, \eta)} |R_n(\theta') - R_n(\theta)| \leq \Delta_n(\varepsilon/2, \eta) + \varepsilon/2 \equiv \Delta'_n(\varepsilon, \eta)$$

with $\Pr(\Delta'_n(\varepsilon, \eta) > \varepsilon) = \Pr(\Delta_n(\varepsilon/2, \eta) > \varepsilon/2) < \eta$. We can show the converse is also true with (A.2) and similar argument. $\square$