1. Introduction.

The consistency of M-estimator replies on two ingredients: the uniform convergence of the sample criterion function to the limit criterion function, and the well-separation of the maximum of the limit function. Under certain conditions, as those in Theorem 2.1 of [NM], the well-separation condition is reduced to the uniqueness of maximum in the limit problem, i.e., an identification problem. We have seen a few examples in the previous lecture for identification: MLE, GMM and QR. There are more examples in Newey and McFadden’s handbook chapter, as well as in Powell’s handbook chapter. We leave these additional examples to the second half of the course.

In this note, we collect some results for establishing the uniform convergence in probability. We will see more results when we talk about empirical processes later.

2. Stochastic equicontinuity

Two key points:

- Definition of stochastic equicontinuity based on the modulus of continuity.
- Alternative formulations of stochastic equicontinuity.

Let \((\Theta, d)\) be a metric space. We first recall the notion of modulus of continuity.

**Definition 1. (Modulus of Continuity)** Let \(f: \Theta \mapsto \mathbb{R}\) be a function. For \(\delta \geq 0\), define

\[
    w(f, \delta) \equiv \sup_{\theta, \theta' \in \Theta, d(\theta, \theta') < \delta} \left| f(\theta) - f(\theta') \right|.
\]

The function \(\delta \mapsto w(f, \delta)\) is called the modulus of continuity of \(f\).

The modulus of continuity quantifies the notion of uniform continuity, as shown below.

**Definition 2. (Uniform Continuity)** A function \(f: \Theta \mapsto \mathbb{R}\) is uniformly continuous iff for every \(\varepsilon > 0\), there exists \(\delta > 0\), such that for every \(\theta, \theta' \in \Theta\) with \(d(\theta, \theta') < \delta\), we have \(\left| f(\theta) - f(\theta') \right| < \varepsilon\).

**Lemma 1. (Modulus of Continuity and Uniform Continuity)** A function \(f: \Theta \mapsto \mathbb{R}\) is uniformly continuous if and only if \(w(f, \delta) \to 0\) as \(\delta \downarrow 0\).

**Proof.** The “if” part. Suppose that \(\lim_{\delta \downarrow 0} w(f, \delta) = 0\). Let \(\varepsilon > 0\). By assumption, there exists \(\delta_0 > 0\) such that \(\delta \leq \delta_0\) implies that \(w(f, \delta) < \varepsilon\). Take any \(\theta, \theta' \in \Theta\) with \(d(\theta, \theta') < \delta_0\). By the definition of \(w(f, \delta)\), we have \(\left| f(\theta) - f(\theta') \right| \leq w(f, \delta_0) < \varepsilon\). Hence, \(f\) is uniformly continuous.

The “only if” part. Suppose that \(f\) is uniformly continuous. Let \(\varepsilon > 0\). By assumption, there exists \(\delta^* > 0\) such that for \(\theta, \theta' \in \Theta\), \(d(\theta, \theta') < \delta^*\) implies that
\[ |f(\theta) - f(\theta')| < \varepsilon/2. \] Hence, 
\[
w(f, \delta^*) = \sup_{\theta, \theta' \in \Theta, d(\theta, \theta') < \delta^*} |f(\theta) - f(\theta')| \leq \varepsilon/2 < \varepsilon.
\]

Hence, for any \( \delta < \delta^* \), we must have \( w(f, \delta) < \varepsilon \). This means that \( \lim_{\delta \downarrow 0} w(f, \delta) = 0 \).

Let \( (Q_n(\theta) : \theta \in \Theta)_{n \in \mathbb{N}} \) be a sequence of random functions defined on \( \Theta \). That is, for each \( \theta \in \Theta \), \( Q_n(\theta) \) is a random variable; for each \( \omega \in \Omega \), the mapping \( \theta \mapsto Q_n(\theta) \) is a deterministic function (the sample path). In particular, for each \( \omega \), we can define the modulus of continuity for the function \( \theta \mapsto Q_n(\theta) \) as above, which we denote by \( w(Q_n, \delta) \). Note that \( w(Q_n, \delta) \) depends on \( \omega \), hence is “random”.

But, this does not mean we can say that \( w(Q_n, \delta) \) is a “random variable”, because \( w(Q_n, \delta) \) may not be \( \mathcal{F} \)-measurable. Again, we shall ignore the measurability issue and simply assume that \( w(Q_n, \delta) \) is a random variable.

**Definition 3. (Stochastic Equicontinuity)** The sequence of random functions \( (Q_n(\theta) : \theta \in \Theta)_{n \in \mathbb{N}} \) is said to be stochastically equicontinuous iff for any \( \varepsilon > 0 \) and \( \eta > 0 \), there exists \( \delta > 0 \) such that
\[
\limsup_{n \to \infty} \mathbb{P}(w(Q_n, \delta) > \varepsilon) < \eta.
\]

In his handbook chapter, Andrews gave two other equivalent definitions for stochastic equicontinuity, see [DA] page 2252. We partially rephrase his statement as the following:

**Theorem 1.** (a) The sequence of random functions \( (Q_n(\theta) : \theta \in \Theta)_{n \in \mathbb{N}} \) is stochastically equicontinuous if and only if for every sequence of constants \( (\delta_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}_+ \) with \( \delta_n \to 0 \), we have \( w(Q_n, \delta_n) \xrightarrow{\mathbb{P}} 0 \).

(b) If \( (Q_n(\theta) : \theta \in \Theta)_{n \in \mathbb{N}} \) is stochastically equicontinuous, then for any two sequences of \( \Theta \)-valued random variables \( (\hat{\theta}_{1n})_{n \in \mathbb{N}} \) and \( (\hat{\theta}_{2n})_{n \in \mathbb{N}} \) with \( d(\hat{\theta}_{1n}, \hat{\theta}_{2n}) \xrightarrow{\mathbb{P}} 0 \), we have \( Q_n(\hat{\theta}_{1n}) - Q_n(\hat{\theta}_{2n}) \xrightarrow{\mathbb{P}} 0 \).

**Remark 1.** In Andrews’ statement, part (b) is an “if and only if” statement.

**Proof.** (Optional) Step 1. In this step, we show the “only if” part of (a). Suppose that \( (Q_n)_{n \in \mathbb{N}} \) is s.e. (i.e. stochastically equicontinuous). Fix \( \varepsilon > 0 \) and let \( \eta > 0 \) be arbitrary. By assumption, there exists \( \delta^* \) such that
\[
\limsup_{n \to \infty} \mathbb{P}(w(Q_n, \delta^*) > \varepsilon) < \eta.
\]
Since $\delta_n \to 0$, for $n$ large enough, $\delta_n < \delta^*$ and hence $w(Q_n, \delta_n) \leq w(Q_n, \delta^*)$. Therefore,

$$\limsup_{n \to \infty} P(w(Q_n, \delta_n) > \varepsilon) < \eta.$$ 

Since $\eta$ is arbitrary, the left-hand-side of the above display must be zero. Hence, $w(Q_n, \delta_n) \not\to 0$.

**Step 2.** In this step, we show the “if” part of (a). Suppose that $w(Q_n, \delta_n) = o_p(1)$ whenever $\delta_n \to 0$. If $(Q_n)_{n \in \mathbb{N}}$ were not s.e., then by definition, there must exist $\varepsilon > 0$ and $\eta > 0$ such that for any $\delta > 0$,

$$\limsup_{n \to \infty} P(w(Q_n, \delta) > \varepsilon) \geq 2\eta > \eta.$$ 

In particular, $\sup_{n \geq 1} P(w(Q_n, 1) > \varepsilon) > \eta$, so there exists $n_1 \geq 1$ such that $P(w(Q_{n_1}, 1) > \varepsilon) > \eta$. Now, suppose that we have obtained a strictly increasing sequence $\{n_m\}_{m \in \{1, \ldots, k-1\}}$ such that $P(w(Q_{n_m}, 1/m) > \varepsilon) > \eta$. By (0.1) with $\delta = 1/k$, we have $\sup_{n > n_{k-1}} P(w(Q_n, 1/k) > \varepsilon) > \eta$. We can thus select $n_k > n_{k-1}$ with $P(w(Q_{n_k}, 1/k) > \varepsilon) > \eta$. In this way, we can select a subsequence $(n_m)_{m \in \mathbb{N}}$ strictly increasing, such that for each $m$,

$$P(w(Q_{n_m}, 1/m) > \varepsilon) > \eta.$$ 

Now, we construct a sequence $(\delta_n)_{n \in \mathbb{N}}$ by setting

$$\delta_n = 1 \text{ when } n \in \{1, \ldots, n_1\},$$
$$\delta_n = 1/2 \text{ when } n \in \{n_1 + 1, \ldots, n_2\},$$
$$\vdots$$
$$\delta_n = 1/m \text{ when } n \in \{n_{m-1} + 1, \ldots, n_m\},$$
$$\vdots$$

By construction $\delta_{n_m} = 1/m$ for each $m \in \mathbb{N}$ and $\lim_{n \to \infty} \delta_n = 0$. But

$$P(w(Q_{n_m}, \delta_{n_m}) > \varepsilon) = P(w(Q_{n_m}, 1/m) > \varepsilon) > \eta.$$ 

This means, $P(w(Q_n, \delta_n) > \varepsilon) \not\to 0$, which contradicts the assumption. By contradiction, we conclude that $(Q_n)_{n \in \mathbb{N}}$ has to be s.e..

**Step 3.** We show (b). Suppose that $(Q_n)_{n \in \mathbb{N}}$ is s.e.. Let $(\hat{\theta}_{1n})_{n \in \mathbb{N}}$ and $(\hat{\theta}_{2n})_{n \in \mathbb{N}}$ be $\Theta$-valued r.v.’s such that $d(\hat{\theta}_{1n}, \hat{\theta}_{2n}) \not\to 0$. Fix $\varepsilon > 0$ and let $\eta > 0$ be arbitrary. Since $(Q_n)_{n \in \mathbb{N}}$ is s.e., there exists $\delta > 0$ such that $\limsup_{n \to \infty} P(w(Q_n, \delta) > \varepsilon) < \eta.$
η. Hence,
\[
\mathbb{P} \left( \left| Q_n \left( \hat{\theta}_{1n} \right) - Q_n \left( \hat{\theta}_{2n} \right) \right| > \varepsilon \right) \\
\leq \mathbb{P} \left( \left| Q_n \left( \hat{\theta}_{1n} \right) - Q_n \left( \hat{\theta}_{2n} \right) \right| > \varepsilon, d \left( \hat{\theta}_{1n}, \hat{\theta}_{2n} \right) \leq \delta / 2 \right) + \mathbb{P} \left( d \left( \hat{\theta}_{1n}, \hat{\theta}_{2n} \right) > \delta / 2 \right).
\]

Taking \( \limsup_{n \to \infty} \) on both sides of the above display and then using \( d \left( \hat{\theta}_{1n}, \hat{\theta}_{2n} \right) = o_p(1) \), we have
\[
\limsup_{n \to \infty} \mathbb{P} \left( \left| Q_n \left( \hat{\theta}_{1n} \right) - Q_n \left( \hat{\theta}_{2n} \right) \right| > \varepsilon \right) \leq \limsup_{n \to \infty} \mathbb{P} (w(Q_n, \delta) > \varepsilon).
\]

Sending \( \delta \downarrow 0 \), we see that the left-hand-side of the above display is 0. Hence, \( Q_n \left( \hat{\theta}_{1n} \right) - Q_n \left( \hat{\theta}_{2n} \right) = o_p(1) \).

3. Uniform convergence and stochastic equicontinuity

We discuss the relationship between uniform convergence and stochastic equicontinuity. The key point is that in order to upgrade pointwise convergence to uniform convergence, we need to establish stochastic equicontinuity.


**Definition 4.** (Totally Bounded Metric Space) A metric space \((\Theta, d)\) is totally bounded if for every \( \varepsilon > 0 \), there exists \( m \in \mathbb{N} \) and \( \{\theta_i\}_{1 \leq i \leq m} \subseteq \Theta \) such that \( \Theta \) is covered by the \( \varepsilon \)-balls centered at \( \theta_i \), \( 1 \leq i \leq m \).

**Remark 2.** A compact metric space is totally bounded. (Royden, Real Analysis, Third Edition, Chapter 7, Proposition 25).

**Theorem 2.** Let \((Q_n(\theta); \theta \in \Theta)_{n \in \mathbb{N}}\) be a sequence of random functions. Consider two statements
(a) \( \sup_{\theta \in \Theta} |Q_n(\theta)| \xrightarrow{p} 0 \)
(b) \( Q_n(\theta) \xrightarrow{p} 0 \) for every \( \theta \in \Theta \) and \((Q_n)_{n \in \mathbb{N}}\) is stochastically equicontinuous.

We have (a) \( \Rightarrow \) (b). If \((\Theta, d)\) is totally bounded, then (b) \( \Rightarrow \) (a).

**Proof.** We first show (a) \( \Rightarrow \) (b). Suppose that \( \sup_{\theta \in \Theta} |Q_n(\theta)| \xrightarrow{p} 0 \). Then it is trivial to see that \( Q_n(\theta) = o_p(1) \) for each \( \theta \in \Theta \). Let \((\delta_n)_{n \in \mathbb{N}}\) satisfy \( \delta_n \to 0 \). Then
\[
w(Q_n, \delta_n) = \sup_{d(\theta, \theta') < \delta_n} |Q_n(\theta) - Q_n(\theta')| \\
\leq 2 \sup_{\theta \in \Theta} |Q_n(\theta)| \xrightarrow{p} 0.
\]

By theorem 1, \((Q_n)_{n \in \mathbb{N}}\) is s.e.
Now, we show that (b)⇒(a) under the additional assumption that \((\Theta,d)\) is totally bounded. Fix \(\varepsilon > 0\) and let \(\eta > 0\) be arbitrary. Since \((Q_n)_{n \in \mathbb{N}}\) is s.e., there exists \(\delta > 0\) such that
\[
\limsup_{n \to \infty} \mathbb{P}(w(Q_n, \delta) > \varepsilon/2) < \eta.
\]
Because \((\Theta,d)\) is totally bounded, \(\Theta\) is covered by finitely many \(\delta\)-balls. That is, there exists \(m \in \mathbb{N}\) and \(\{t_i\}_{1 \leq i \leq m}\) such that for any \(\theta \in \Theta\), there exists \(i(\theta) \in \{1, \ldots, m\}\) such that \(d(\theta, t_i(\theta)) < \delta\). Hence,
\[
\sup_{\theta \in \Theta} |Q_n(\theta)| \leq \sup_{\theta \in \Theta} |Q_n(\theta) - Q_n(t_i(\theta))| + \sup_{\theta \in \Theta} |Q_n(t_i(\theta))| \\
\leq w(Q_n, \delta) + \max_{1 \leq i \leq m} |Q_n(t_i)|.
\]
Hence,
\[
\limsup_{n \to \infty} \mathbb{P}\left(\sup_{\theta \in \Theta} |Q_n(\theta)| > \varepsilon\right) \\
\leq \limsup_{n \to \infty} \mathbb{P}\left(w(Q_n, \delta) > \varepsilon/2\right) + \limsup_{n \to \infty} \mathbb{P}\left(\max_{1 \leq i \leq m} |Q_n(t_i)| > \varepsilon/2\right).
\]
But \(Q_n(t_i) = o_p(1)\) also implies that \(\max_{1 \leq i \leq m} |Q_n(t_i)| = o_p(1)\). Hence the second term on the majorant side of the above display is 0. Hence,
\[
\limsup_{n \to \infty} \mathbb{P}\left(\sup_{\theta \in \Theta} |Q_n(\theta)| > \varepsilon\right) < \eta.
\]
Since \(\eta\) is arbitrary, we have \(\mathbb{P}(\sup_{\theta \in \Theta} |Q_n(\theta)| > \varepsilon) \to 0\). That is, \(\sup_{\theta \in \Theta} |Q_n(\theta)| \overset{p}{\to} 0\).

**Remark 3.** The limiting function here is supposed to be 0. But this is only a normalization. If we are interested in \(Q_n(\theta) \to Q(\theta)\), we can always define \(\hat{Q}_n(\theta) = Q_n(\theta) - Q(\theta)\), and then apply the theorem to \((\hat{Q}_n)_{n \in \mathbb{N}}\).

**Problem 1.** Let \((\Theta,d)\) be a metric space. Let \((G_n)_{n \in \mathbb{N}}\) and \((Q_n)_{n \in \mathbb{N}}\) be two sequences of random functions, both are stochastically equicontinuous. Then \((G_n + Q_n)_{n \in \mathbb{N}}\) is also s.e.

**Problem 2.** Let \((\Theta,d)\) be a metric space and let \(Q\) be a deterministic function on \(\Theta\). Then the sequence \((Q_n)_{n \in \mathbb{N}}\) defined by \(Q_n = Q\) for every \(n \in \mathbb{N}\) is s.e. if and only if \(Q\) is uniformly continuous.

4. Establishing stochastic equicontinuity

4.1. Lipschitz-type conditions.
Theorem 3. ([JD] Theorem 21.10, Andrews 1992) Suppose there exists \( N \in \mathbb{N} \) such that almost surely,
\[
|Q_n(\theta) - Q_n(\theta')| \leq B_n h(d(\theta, \theta'))
\]
holds for all \( \theta, \theta' \in \Theta \) and \( n \geq N \), where \( h \) is a deterministic function and \( h(x) \downarrow 0 \) as \( x \downarrow 0 \), and \( B_n = O_p(1) \). Then \( (Q_n)_{n \in \mathbb{N}} \) is s.e.

**Proof.** Let \( \delta_n \to 0 \). Then for \( n \) sufficiently large,
\[
w(Q_n, \delta_n) = \sup_{\theta, \theta' \in \Theta, d(\theta, \theta') < \delta_n} |Q_n(\theta) - Q_n(\theta')| \\
\leq B_n h(\delta_n) = O_p(1) = o_p(1).
\]
Hence, by Theorem 1, we conclude that \( (Q_n)_{n \in \mathbb{N}} \) is s.e.. \( \Box \)

**Remark 4.** An important special case of this result is the one with \( h(x) = x \). Then the displayed condition of the theorem is simply \( Q_n(\theta) - Q_n(\theta') \leq B_n d(\theta, \theta') \). That is, \( Q_n \) is Lipschitz in its index with stochastically bounded coefficient \( B_n \).

**Remark 5.** Typically, the function \( Q_n(\theta) \) has the form \( Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} q(X_t, \theta) \). The good thing of the theorem is that it does not require \( X_t \) to be independent and/or identical. The bad thing is that the Lipschitz condition on \( Q_n(\theta) \) is often transferred to \( q(X_t, \theta) \), which is too strong for some applications.

**Remark 6.** \( Q_n(\theta) \) does not have to be of the form \( \frac{1}{n} \sum_{t=1}^{n} q(X_t, \theta) \). For example, one may take \( Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} q(X_t, \hat{\gamma}_n, \theta) \), where \( \hat{\gamma}_n \) is some preliminary estimator. The theorem might still be useful when \( q(\cdot) \) is Lipschitz in \( \theta \).

The displayed condition in Theorem 3 can be easily checked if \( Q_n \) is differentiable. To be precise, we have the following:

**Corollary 1.** Let \( q \in \mathbb{N} \). Let \( \Theta \subseteq \mathbb{R}^d \) be convex. Let \( (Q_n)_{n \in \mathbb{N}} \) be a sequence of real-valued random functions on \( \Theta \). Suppose that \( Q_n \) is differentiable and \( \sup_{\theta \in \Theta} \| \nabla_\theta Q_n(\theta) \| = O_p(1) \). Then \( (Q_n)_{n \in \mathbb{N}} \) is s.e..

**Proof.** By the mean value theorem, we have, for \( \theta, \theta' \in \Theta \) and some \( \tilde{\theta} \) in between,
\[
Q_n(\theta) - Q_n(\theta') = (\theta - \theta')^\top \nabla_\theta Q_n(\tilde{\theta}).
\]
Hence,
\[
|Q_n(\theta) - Q_n(\theta')| \leq \| \theta - \theta' \| \| \nabla_\theta Q(\tilde{\theta}) \| \\
\leq \sup_{\theta \in \Theta} \| \nabla_\theta Q_n(\theta) \| \| \theta - \theta' \|.
\]
The claim then follows from Theorem 3. \( \Box \)
Below, we illustrate the application of Theorem 3 and Corollary 1 via some simple examples. We first recall an important inequality for studying polynomials:

**Lemma 2.** (Loève’s Cr inequality, [JD] (9.63)) Let $r > 0$, $m \in \mathbb{N}$ and $\{x_i\}_{1 \leq i \leq m} \subseteq \mathbb{R}$. Then $|\sum_{i=1}^{m} x_i|^r \leq c_r \sum_{i=1}^{m} |x_i|^r$, where

$$c_r = \begin{cases} 
1 & \text{when } r \leq 1 \\
 m^{-r} & \text{when } r \geq 1
\end{cases}.$$

**Example 1.** (Linear model under polynomial loss, the smooth case) Let $\Theta \subseteq \mathbb{R}$. Let $Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} |Y_t - X_t \theta|^p$ for some $p > 1$. When $p = 2$, we have OLS. Since $p > 1$, $Q_n(\theta)$ is differentiable. It is easily seen that $|Q_n'(\theta)| \leq \frac{1}{n} \sum_{t=1}^{n} p |Y_t - X_t \theta|^{p-1} |X_t|$. Hence,

$$\sup_{\theta \in \Theta} |Q_n'(\theta)| \leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} p |Y_t - X_t \theta|^{p-1} |X_t| \leq K_p \left( \frac{1}{n} \sum_{t=1}^{n} |Y_t|^{p-1} |X_t| + \left( \sup_{\theta \in \Theta} |\theta|^p \right) \frac{1}{n} \sum_{t=1}^{n} |X_t|^p \right) \leq K_p \left( \frac{1}{n} \sum_{t=1}^{n} |Y_t|^p \right)^{\frac{p-1}{p}} \left( \frac{1}{n} \sum_{t=1}^{n} |X_t|^p \right)^{\frac{1}{p}} + K_p \left( \sup_{\theta \in \Theta} |\theta|^p \right) \left( \frac{1}{n} \sum_{t=1}^{n} |X_t|^p \right),$$

where $K_p$ is some constant only depending on $p$, the second inequality follows from the Cr-inequality, the third inequality follows from Hölder’s inequality. If $\Theta$ is bounded and both $\frac{1}{n} \sum_{t=1}^{n} |Y_t|^p$ and $\frac{1}{n} \sum_{t=1}^{n} |X_t|^p$ are $O_p(1)$, then $\sup_{\theta \in \Theta} |Q_n'(\theta)| = O_p(1)$. The previous corollary then implies that $(Q_n)_{n \in \mathbb{N}}$ is s.e..

**Example 2.** (Linear model under polynomial loss, the nonsmooth case) Consider the same setting as the previous example, but now take $0 < p \leq 1$. In this case, $Q_n(\theta)$ is no longer differentiable. Nevertheless, by the Cr-inequality, it is easy to see that

$$x, y \in \mathbb{R} \Rightarrow ||x + y|^p - |x|^p| \leq |y|^p.$$ 

Hence,

$$|Q_n(\theta) - Q_n(\theta')| \leq \frac{1}{n} \sum_{t=1}^{n} ||Y_t - X_t \theta|^p - |Y_t - X_t \theta'|^p| \leq \frac{1}{n} \sum_{t=1}^{n} |X_t|^p |\theta - \theta'|^p.$$

If we assume that $\frac{1}{n} \sum_{t=1}^{n} |X_t|^p = O_p(1)$, then taking $h(x) = |x|^p$ and $B_n = \frac{1}{n} \sum_{t=1}^{n} |X_t|^p$ in Theorem 3, we conclude that $(Q_n)_{n \in \mathbb{N}}$ is s.e.
Problem 3. (A simple MLE) Let \((X_t)_{1 \leq t \leq n}\) be iid \(N(\mu, v)\) distributed random variables. Let the parameter space \(\Theta = \{ (\mu, v) : |\mu| \leq \bar{\mu}, v \in [v, \hat{v}] \}\) for some finite \(\bar{\mu}\) and \(\hat{v}\) and \(v > 0\).

(a) Write down the likelihood function \(Q_n\).
(b) Show that \(Q_n\) converges in probability uniformly to some function \(Q\).
(c) Show that the MLE is consistent.

Problem 4. (Median Regression) Let \(Y_t = X_t^T \theta_0 + \varepsilon\) where \(\theta_0 \in \Theta \subseteq \mathbb{R}^q\) for some \(q \in \mathbb{N}\) and compact \(\Theta\). Suppose that median \((\varepsilon | X) = 0\). Let \(Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} |Y_t - X_t \theta|\) and \(\hat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\theta)\).

(a) Show that \(Q_n\) converges in probability uniformly to some deterministic function \(Q\). Design conditions for this to hold. What is your \(Q\) function?
(b) Show that \(\hat{\theta}_n \xrightarrow{P} \theta_0\). Be clear about what you assume, and then prove your claim.

4.2. Results for sample averages. We now consider the uniform convergence of \(Q_n(\theta)\) with the following form

\[ Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} q(X_t, \theta). \]

The key here is that we will allow \(q(X_t, \theta)\) to be discontinuous in \(\theta\); this generalization is crucial for many semiparametric calculations, especially those involves indicator functions.

Theorem 4. (Lemma 2.4 of [NM], Tauchen 1985 Lemma 1) Let \((X_t)_{t \in \mathbb{N}}\) be IID \(\mathbb{R}^m\)-valued random variables and \(Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} q(X_t, \theta)\) for \(\theta \in \Theta\) and some (separable) function \(q : \mathbb{R}^m \times \Theta \mapsto \mathbb{R}^k\). Suppose that

(a) \(\Theta\) is compact;
(b) for each \(\theta \in \Theta\), the function \(x \mapsto q(x, \theta)\) is measurable;
(c) there exists some function \(b : \mathbb{R}^m \mapsto \mathbb{R}_+\) such that \(\sup_{\theta \in \Theta} \| q(x, \theta) \| \leq b(x)\), and \(\mathbb{E} [b(X_t)] < \infty\);
(d) for each \(\theta \in \Theta\), there exists \(A_\theta \subseteq \mathbb{R}^m\) such that \(\mathbb{P}(X_t \in A_\theta) = 0\) and for each \(x \in A_\theta^c\) the function \(q(x, \cdot)\) is continuous at \(\theta\).

Then \(Q_n(\theta) \xrightarrow{P} Q(\theta) \equiv \mathbb{E}[q(X_t, \theta)]\) uniformly in \(\theta \in \Theta\).

Remark 7. The beauty of this result is that it allows \(q(X_t, \theta)\) to be discontinuous in \(\theta\). The usefulness relies on the fact that in (d) the null set \(A_\theta\) may depend on \(\theta\).

Remark 8. This result exploits the special feature that \(Q_n(\theta)\) is a sample average. Uniform convergence of such \(Q_n\) is called a “Glivenko-Cantelli” theorem, which we will explore further when we talk about empirical processes.
Example 3. (Maximum Score, [NM] 2.8.1, Manski 1975) Let $Z_t = (Y_t, X_t^\top)^\top$ and $q(Z_t, \theta) = \left| Y_t - 1 \{X_t^\top \theta > 0\} \right|$. Let $\Theta = \{\theta \subseteq \mathbb{R}^m : \|\theta\| = 1\}$. Note that $\sup_{\theta \in \Theta} |q(Z_t, \theta)| \leq |Y_t| + 1$. In the binary choice model, $Y_t$ can only take values in $\{0, 1\}$, hence bounded. Condition (c) is thus verified with $b(x) = 2$. For (d), note that $\theta \mapsto q(Z_t, \theta)$ is discontinuous only when $X_t^\top \theta = 0$. If the distribution of $X_t$ put zero mass on each hyperplane through the origin, then $\theta \mapsto q(Z_t, \theta)$ is continuous at each $\theta \in \Theta$ with probability 1; the null set depends on $\theta$. 