Extremum Estimation

An extremum estimator is any estimator defined as

$$\hat{\theta} = \arg\min Q_n(w_1, \ldots, w_n; \theta)$$

for some parameter θ in parameter space ④
and $w_1, \ldots, w_n$ is the sample

- $Q_n(\cdot)$ = criterion / objective function
  - ↳ indexed by n because it must depend on your sample
  - ↳ $Q(\cdot)$ = population analog of $Q_n(\cdot)$

Note: Under extremum estimation, we make no assumptions about $Q_n(\cdot)$ aside from that fact that $Q_n(\cdot)$ must be a function of the data $(w_1, \ldots, w_n)$ and θ.

- $Q_n(\cdot)$ does not necessarily need to be in some quadratic form (like GMM)
  - ↳ encompasses nonlinear objective functions that may require calculation through numerical methods
  - ↳ ie: may not have closed form solution

Aside from the assumptions later on, Extremum Estimation only requires that we minimize/maximize an objective function.

	GMM	Extremum Estimation
Obj Function	$J_n(\beta) = \arg\min \bar{g}_n(\beta)' W \bar{g}_n(\beta)$ where $\bar{g}_n(\beta)$ = sample moment conditions ↳ Note: Explicit weighted distance structure in the objective function	$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min}\, Q_n(w_1, \ldots, w_n; \theta)$ ↳ could be any form (as long as it meets assumptions) ↳ not necessarily moment conditions
Goal	Find $\hat{\beta}$ that sets my moment conditions $\bar{g}_n(\beta)$ as close to 0 as possible	Find θ to minimize objective function

Consistency

(Note: Assuming GMM linear)

Clearly,

$$\bar{g}_n(\beta) \xrightarrow{p} \bar{g}(\beta) \quad \text{by WLLN}$$

We also choose

$$\hat{W} \xrightarrow{p} W$$

Hence,

$$\hat{\beta} \xrightarrow{p} \beta \quad \text{by CMT, Slutsky}$$

since

for unique solution

- identification (ie: $(X'Z)$ full rank)
- convergence (ie: LLN, CLT, Slutskys)
- closed form solution (easy to prove consistency) if parts conv → whole thing conv

If $Q_n(\theta) \xrightarrow{p} Q(\theta)$,

does $\hat{\theta} \xrightarrow{p} \theta$?

↳ what else do I need?

- Identification
  ↳ how do I know my θ is unique?
- can't just use LLN/CLT bc only gives you pointwise conv.

  ie: $\exists \tilde{\theta}$ where $g_n(\tilde{\theta}) \xrightarrow{p} g(\theta)$
  but for other values of θ,
  $g_n(\theta) \xrightarrow{p} g(\theta)$ may not be true

Examples of Extremum Estimation

- Nonlinear Least Squares

$$
\begin{aligned}
Q(\theta) &= \mathbb{E}(Y - m(X,\theta))^2 \\
Q_n(\theta) &= \mathbb{E}_n(Y - m(X,\theta))^2.
\end{aligned}
$$

- GMM

$$
\begin{aligned}
Q(\theta) &= \mathbb{E}g(\theta)'\boldsymbol{W}\mathbb{E}g(\theta) \\
Q_n(\theta) &= \mathbb{E}_n g(\theta)'\hat{\boldsymbol{W}}\mathbb{E}_n g(\theta).
\end{aligned}
$$

- Maximum Likelihood Estimation

$$
\begin{aligned}
Q(\theta) &= \mathbb{E}\ell(W;\theta) \\
Q_n(\theta) &= \mathbb{E}_n\ell(W;\theta).
\end{aligned}
$$

M - Estimation

Special case of Extremum Estimation where your criterion function is defined as:

$$Q(\theta) = \mathbb{E}m(W;\theta)$$
$$Q_n(\theta) = \mathbb{E}_n m(W;\theta)$$

where $m(\cdot)$ is some known real-valued function

Note: M-estimation only requires that your criterion function be in the form of a sample avg.

Ex: If we take the log of the likelihood, we end up solving for

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \mathbb{E}_n m(W, \theta)$$
$$= \arg\max_{\theta \in \Theta} \underbrace{\frac{1}{n} \sum \log f(Y_i \mid X_i, \theta)}_{\text{log-likelihood}}$$

Consistency

Note: Since there are multiple versions of consistency theorems for extremum estimation, I will give them unofficial names

[Consistency Thm (A)]

Theorem

Assume:

1. The sample criterion uniformly consistently estimates the population criterion:

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

2. θ_0 is a unique and well-separated global minimum of $Q(\cdot)$:

$$\forall \epsilon > 0 \exists \delta > 0 : Q^{\epsilon} \equiv \inf_{\theta \in \Theta : \|\theta - \theta_0\| \geq \epsilon} Q(\theta) \geq Q(\theta_0) + \delta.$$

Then $\hat{\theta} \xrightarrow{p} \theta_0$.

To illustrate why we need these assumptions to conclude consistency, we will

- setup the motivation for these assumptions
- define the terms
- go back and see why these assumptions fix our issues

[Consistency Thm Ⓐ : Uniform Convergence]

(Hansen p 781)

(ex: $\hat{\beta} \xrightarrow{P} \beta$)

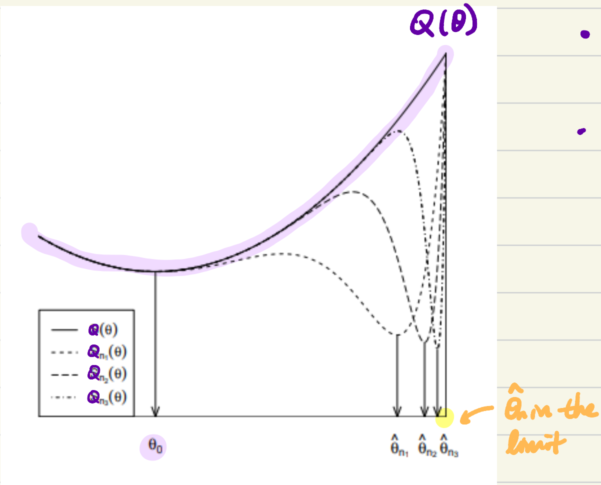
In linear estimators, we apply WLLN to conclude consistency

__Issue__: Nonlinear estimators may not have a closed form solution
↳ cannot easily conclude consistency using WLLN

However, we have the sample criterion function $Q_n(\theta)$, which is a sample average

$\Rightarrow$ Hence, $Q_n(\theta) \xrightarrow{P} Q(\theta)$ by WLLN

__BUT__, if $Q_n(\theta) \xrightarrow{P} Q(\theta) \overset{?}{\Longrightarrow} \hat{\theta} \xrightarrow{P} \theta$?



- True function $Q(\theta)$ is a curve
- However, we can find a $Q_n(\theta)$ where:
  - $Q_n(\theta) \to Q(\theta)$
  - $Q_n(\theta)$ has a sharp dip at the limit of the parameter space

Hence, we have:

$Q_n(\theta) \xrightarrow{P} Q(\theta)$ but $\hat{\theta}_n \not\xrightarrow{P} \theta_0$

⊛ Uniform Convergence

Definition 22.1 $S_n(\theta)$ **converges in probability** to $S(\theta)$ **uniformly** over $\theta \in \Theta$ if

$$\sup_{\theta \in \Theta} |S_n(\theta) - S(\theta)| \xrightarrow[p]{} 0$$

as $n \to \infty$.

- convergence happens uniformly across the entire parameter space
- can't have regions of θ where convergence is slow

In other words, for n sufficiently large, the sample criterion function satisfies

$$\forall \varepsilon > 0, \lim_{n \to \infty} P\left(\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| < \varepsilon\right) = 1$$

- For large n, behavior of $Q_n(\theta)$ is bounded around $[Q(\theta) - \varepsilon, Q(\theta) + \varepsilon]$

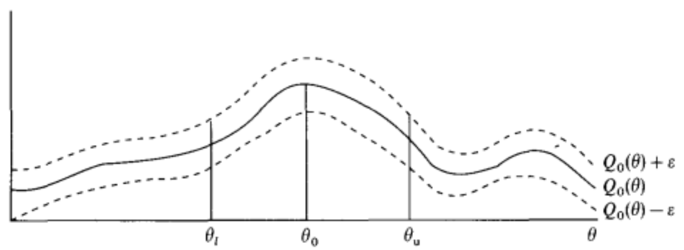


Figure 2.

⊛ Well-Separated

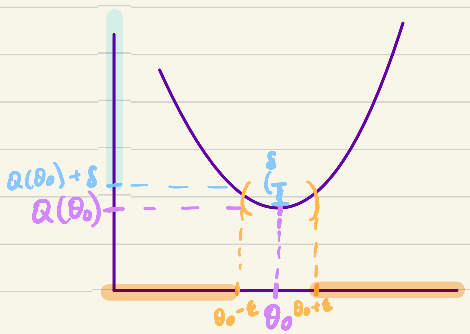
Obviously, our true θ_0 must be unique.

However, if we have multiple θ-values near θ_0 that almost minimize $Q_n(\theta)$, then we might have an issue with slow convergence

Ex:

Definition: θ_0 is well-separated if

$$\forall \epsilon > 0 \exists \delta > 0 : Q^{\epsilon} \equiv \inf_{\theta \in \Theta : \|\theta - \theta_0\| \geq \epsilon} Q(\theta) \geq Q(\theta_0) + \delta.$$



ie: Choose $\theta \in \Theta$, $\|\theta - \theta_0\| \geq \epsilon$ (not in $B_\epsilon(\theta_0)$) that make $Q(\theta)$ as small as possible, then my $Q(\theta)$ is still larger than $Q(\theta_0) + \delta$

Consistency Thm (A) : Hansen's Version

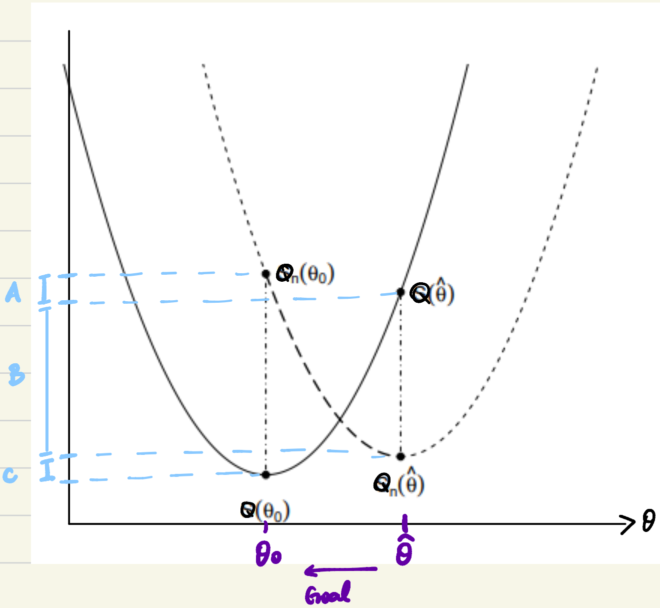
Theorem 22.1 $\widehat{\theta} \underset{p}{\longrightarrow} \theta_0$ as $n \to \infty$ if

1. $S_n(\theta)$ converges in probability to $S(\theta)$ uniformly over $\theta \in \Theta$.
2. θ_0 uniquely minimizes $S(\theta)$ in the sense that for all $\epsilon > 0$,

$$\inf_{\theta:\|\theta-\theta_0\|\geq\epsilon} S(\theta) > S(\theta_0).$$

Goal: Prove $\hat{\theta} \xrightarrow{P} \theta_0$

Issue: $Q_n \to Q$ in order for sample min $\to$ true min



Proof Structure:

1) $Q(\hat{\theta}) \xrightarrow{P} Q(\theta_0)$
   - ↳ minimize vertical dist
   - ↳ Show $B+C \to 0$

2) $\hat{\theta} \xrightarrow{P} \theta_0$
   - ↳ minimize horizontal dist

Consistency Thm (A) : Hansen's Version

Key Points to Remember:

- (Population) θ_0 minimizes $Q(\theta)$
- (Sample) $\hat{\theta}$ minimizes $Q_n(\theta)$ (*)
- $\underbrace{\theta_0}_{\text{(pop)}}$ may not minimize $\underbrace{Q_n(\theta)}_{\text{sample}}$

Proof 1) Show $Q(\hat{\theta}) \xrightarrow{P} Q(\theta_0)$

Since θ_0 minimizes, $Q(\theta)$ then $\hat{Q}(\theta_0) \leq Q(\theta)$

$$\Rightarrow 0 \leq Q(\hat{\theta}) - Q(\theta_0)$$

$$= \underbrace{Q(\hat{\theta}) \overset{B}{-} Q_n(\hat{\theta})}_{\substack{\text{bounded by} \\ \sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{P} 0}} + \underbrace{\overset{A+B+C}{Q_n(\theta_0) - Q(\theta_0)}}_{} + \underbrace{\overset{-B-A}{Q_n(\hat{\theta}) - Q_n(\theta_0)}}_{\substack{Q_n(\hat{\theta}) \leq Q_n(\theta_0) \\ \Rightarrow Q_n(\hat{\theta}) - Q_n(\theta_0) \leq 0 \\ \text{by (*)}}}$$

$$\leq 2 \underbrace{\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)|}_{\substack{\text{uniform convergence} \xrightarrow{P} 0 \\ \text{(by assumption)}}} \xrightarrow{P} 0$$

Hence, $Q(\hat{\theta}) \xrightarrow{P} Q(\theta)$

2) Show $\hat{\theta} \xrightarrow{P} \theta_0$

Let $\varepsilon > 0$ and assume we have a unique minimum

Then $\exists \delta > 0$ s.t

DEF of well-separated

$$\|\theta_0 - \theta\| > \varepsilon \implies Q(\theta) - Q(\theta_0) \geq \delta$$

Since the above statement is for *any* $\theta \in \Theta$, then

$$\|\theta_0 - \hat{\theta}\| > \varepsilon \implies Q(\hat{\theta}) - Q(\theta_0) \geq \delta$$

$$\circledast \implies P(\|\theta_0 - \hat{\theta}\| > \varepsilon) \leq \underbrace{P(Q(\hat{\theta}) - Q(\theta_0) \geq \delta) \xrightarrow{P} 0}_{\text{Part 1: } Q(\hat{\theta}) \xrightarrow{P} Q(\theta_0)}$$

Hence,

$$P(\|\theta_0 - \hat{\theta}\| > \varepsilon) \xrightarrow{P} 0$$

$$\implies \hat{\theta} \xrightarrow{P} \theta_0$$

Note: If $A \implies B$, then $P(A) \leq P(B)$.

$A \implies B$ implies $A \cap B = A$.

Then by partition theorem,

$$P(B) = P(B \cap A) + P(B \cap A^c)$$

$$= P(A) + \underbrace{P(B \cap A)}_{\geq 0} \geq P(A)$$

Proof of Consistency Thm Ⓐ

(In class theorem)

Theorem

Assume:

1. The sample criterion uniformly consistently estimates the population criterion:

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

2. θ_0 is a unique and well-separated global minimum of $Q(\cdot)$:

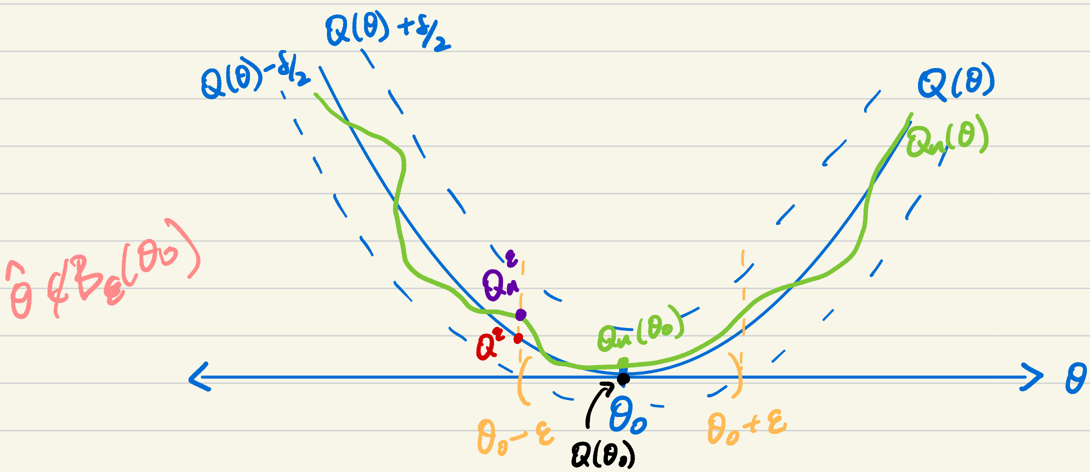
$$\forall\epsilon > 0 \exists\delta > 0 : Q^\epsilon \equiv \inf_{\theta\in\Theta:\|\theta-\theta_0\|\geq\epsilon} Q(\theta) \geq Q(\theta_0) + \delta.$$

Then $\hat\theta \xrightarrow{p} \theta_0$.

Proof

Fix $\epsilon > 0$ and define $Q_n^\epsilon \equiv \inf_{\theta\in\Theta:\|\theta-\theta_0\|\geq\epsilon} Q_n(\theta)$, then

$$\begin{aligned}
& \Pr(\|\hat\theta - \theta_0\| > \epsilon) \\
\leq\; & \Pr\big(Q_n^\epsilon \leq Q_n(\theta_0)\big) \\
=\; & 1 - \Pr\big(Q_n^\epsilon > Q_n(\theta_0)\big) \\
\leq\; & 1 - \Pr\big(Q_n^\epsilon > Q_\epsilon - \delta/2, Q_n(\theta_0) < Q(\theta_0) + \delta/2\big) \\
\to\; & 0,
\end{aligned}$$



Recall,

$$\forall \epsilon > 0 \exists \delta > 0 : Q^{\epsilon} \equiv \inf_{\theta \in \Theta : \|\theta - \theta_0\| \geq \epsilon} Q(\theta) \geq Q(\theta_0) + \delta.$$

$\Rightarrow$ I choose θ s.t θ is outside of my $B_\varepsilon(\theta_0)$ and minimize my value of $Q(\theta)$

Before we begin this proof, let's map out our existing relationships

a) $Q^\varepsilon \geq Q(\theta_0) + \delta$ def of well-separated

$Q^\varepsilon - \frac{\delta}{2} \geq Q(\theta_0) + \frac{\delta}{2}$

b) $Q_n(\theta_0) < Q_n^\varepsilon$ def of Q_n^ε (if we know that $\hat{\theta} \in B_\varepsilon(\theta_0)$) $\Rightarrow$ $\|\hat{\theta} - \theta_0| > \varepsilon$ then $Q_n(\theta_0) > Q_n^\varepsilon$

From our graph (by construction)

c) $Q_n(\theta_0)$, $Q(\theta_0)$ at most $\frac{\delta}{2}$ away

$\Rightarrow Q(\theta_0) + \frac{\delta}{2} \geq Q_n(\theta_0)$

d) Q_n^ε, Q^ε at most $\frac{\delta}{2}$ away

$\Rightarrow \underline{Q_n^\varepsilon + \frac{\delta}{2} > Q^\varepsilon}$ / $Q^\varepsilon + \frac{\delta}{2} > Q_n^\varepsilon$

Since $Q_n^\varepsilon > Q_n(\theta_0)$

$$\Rightarrow Q_n^\varepsilon \overset{d}{>} Q^\varepsilon - \frac{\delta}{2} \overset{a}{\geq} Q(\theta_0) + \frac{\delta}{2} \overset{c}{>} Q_n(\theta_0)$$

Fix $\epsilon > 0$ and define $Q_n^\epsilon \equiv \inf_{\theta\in\Theta:\|\theta-\theta_0\|\geq\epsilon} Q_n(\theta)$, then

$$\begin{aligned}
& \Pr(\|\hat{\theta}-\theta_0\| > \epsilon) \\
\leq\ & \Pr\big(Q_n^\epsilon \leq Q_n(\theta_0)\big) \\
=\ & 1-\Pr\big(Q_n^\epsilon > Q_n(\theta_0)\big) \\
\leq\ & 1-\Pr\big(Q_n^\epsilon > Q_\epsilon - \delta/2, Q_n(\theta_0) < Q(\theta_0)+\delta/2\big) \\
\to\ & 0,
\end{aligned}$$

① ②

① $\|\hat{\theta}-\theta_0\| > \varepsilon \implies Q_n^\varepsilon \leq Q_n(\theta_0)$

$P(A)$ $P(B)$

② Work from above

$$\text{★}\quad Q_n^\varepsilon \overset{\text{ⓔ}}{>} Q^\varepsilon - \frac{\delta}{2} \overset{\text{ⓐ}}{\geq} Q(\theta_0)+\frac{\delta}{2} \overset{\text{ⓓ}}{>} Q_n(\theta_0)$$

well-separated (existing assumption)

$$1-P(\text{★}) \leq 1-P\Big(\underbrace{\big(Q_n^\varepsilon > Q^\varepsilon - \tfrac{\delta}{2}\big)}_{\text{uniform convg}} \cap \underbrace{\big(Q(\theta_0)+\tfrac{\delta}{2} > Q_n(\theta_0)\big)}_{\text{pointwise convg}}\Big) \xrightarrow{P} 1$$

split events

$\longrightarrow 0$

Consistency Thm (B) (Prep)

Issue: Some of these conditions are hard to check in real life?

Can I create lower-level assumptions such that I can guarantee a well-separated minimum?

Theorem

Assume that:

1. $Q(\cdot)$ is continuous,
2. Θ is compact,
3. θ_0 uniquely minimizes $Q(\theta)$.

Then θ_0 is a well-separated minimum.

And for uniform convergence:

Theorem

Assume that:

1. $\hat{\theta}$ is an m-estimator,
2. The data are i.i.d. realizations of W,
3. $m(W;\theta)$ is a.s. continuous in θ,
4. $|m(W;\theta)| \leq G(W)$ for some function G s.t. $\mathbb{E}G(W) < \infty$,
5. Θ is compact.

Then $Q_n(\cdot)$ converges to $Q(\cdot)$ uniformly.

Proof

This is the Uniform Law of Large Numbers.

Recall:

$$\begin{aligned} Q(\theta) &= \mathbb{E}m(W;\theta) \\ Q_n(\theta) &= \mathbb{E}_n m(W;\theta) \end{aligned}$$

Consistency Thm (B)

Consolidated Theorem

Assume that:

1. $Q(\cdot)$ is continuous,
2. Θ is compact,
3. θ_0 uniquely minimizes $Q(\theta)$,
4. $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| \xrightarrow{p} 0$.

Then $\hat{\theta} \xrightarrow{p} \theta_0$.

Consistency Thm ©

Consistency for Convex $Q(\cdot)$

Assume that:

1. Θ is convex,
2. $\theta_0 \in \operatorname{int} \Theta$,
3. θ_0 uniquely minimizes $Q(\theta)$,
4. $Q_n(\cdot)$ is convex,
5. $|Q_n(\theta) - Q(\theta)| \xrightarrow{p} 0, \forall \theta \in \Theta$.

Then $\hat{\theta} \xrightarrow{p} \theta_0$.