Chapter 1

Empirical Process Theory

1.1 Motivation

OLS and 2SLS have closed-form solutions. In general, however, GMM, NLS and MLE estimators cannot be expressed in closed-form. New asymptotic apparatus are needed for them.

Example 1.1. The maximum likelihood estimator is

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \log f\left(X_i, \theta\right) = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} -\log f\left(X_i, \theta\right).$$

We will need a *uniform* law of large numbers to establish the consistency of $\widehat{\theta}$. Pointwise convergence at every $\theta \in \Theta$ is not sufficient. [to add a diagram to show pointwise convergence does not guarantee uniform convergence.]

Let $g\left(x, \theta\right) : \mathcal{X} \times \Theta \mapsto \mathbb{R}^K$. Here x is the realized value of a random variable, and θ is the parameter. The statistical model is indexed by $\theta \in \Theta$. To keep it simple, we consider Θ a subset of a finite-dimensional Euclidean space $\mathbb{R}^D$. In Example 1.1 $g\left(X_i\left(\omega\right), \theta\right) = -\log f\left(X_i\left(\omega\right), \theta\right)$.

We are interested in large sample behaviors of two objects: the *sample average*

$$\bar{g}_n\left(\theta\right) = \frac{1}{n}\sum_{i=1}^{n} g\left(X_i, \theta\right) = \frac{1}{n}\sum_{i=1}^{n} g\left(X_i\left(\omega\right), \theta\right)$$

and the *normalized average*

$$\nu_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left(g\left(X_i, \theta\right) - E\left[g\left(X_i, \theta\right)\right]\right).$$

Some books call this form of ν_n *empirical process.*

Throughout this chapter, we discuss iid $\left(X_i\right)_{i=1}^n$ for simplicity. Without loss of generality, the marginal distribution of X_i can be represented by X.

Example 1.2. A classical statistical example that motivates the Empirical Cumulative Distribution Function (ECDF)

$$F_n\left(\theta\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left(X_i \leq \theta\right)$$

for $\theta \in \mathbb{R}$. The pointwise LLN and CLT yield

$$F_n(\theta) \xrightarrow{p} F(\theta) = P(X_1 \leq \theta)$$
$$\sqrt{n}(F_n(\theta) - F(\theta)) \xrightarrow{d} N(0, F(\theta)(1 - F(\theta))).$$

Can we strengthen the convergence to hold uniformly over $\theta \in \mathbb{R}$? What is the asymptotic distribution of $\sqrt{n}(F_n(\theta) - F(\theta))$ as a random function indexed by θ?

1.2 Complexity of a Family of Functions

We first work with uniform convergence. To generalize pointwise convergence to uniform convergence, it is essential to study the complexity of the class of functions

$$\mathcal{G} := \{g(\cdot, \theta) : \theta \in \Theta\}$$

under the probability distribution Q that generates the random variables X.

For a generic function $h(x)$, define the L_r-norm as

$$\|h\|_{Q,r} = (E_Q[\|h(X)\|^r])^{1/r} = \left(\int \|h(X)\|^r \, \mathrm{d}Q\right)^{1/r}.$$

A function $G(x)$ is called an **envelope function** of $\mathcal{G}$ if $\sup_{\theta\in\Theta} \|g(x,\theta)\| \leq G(x)$ for every $x \in \mathcal{X}$. The envelope function works as the dominating function in the celebrated dominated convergence theorem.

The **distance** between two functions $g(\cdot, \theta_1)$ and $g(\cdot, \theta_2)$ is defined as $\|g(\cdot, \theta_1) - g(\cdot, \theta_2)\|_{Q,r}$. Given the distance, we define an ε-neighborhood centered around $g(\cdot, \theta)$:

$$\mathcal{N}_{\varepsilon,r,Q}(g(\cdot,\theta)) := \left\{h \in \mathcal{G} : \|h(\cdot) - g(\cdot,\theta)\|_{Q,r} \leq \varepsilon\right\}.$$

The **covering number**, denoted as $N_r(\varepsilon, Q)$, is the smallest number of points m such that $\mathcal{G} \subseteq \bigcup_{\ell=1}^{m} \mathcal{N}_{\varepsilon,r,Q}(g(\cdot,\theta_\ell))$. In other words, those m neighborhoods cover the set of functions $\mathcal{G}$.

Remark 1.1. It is straightforward to generalize from pointwise convergence to uniform convergence if Θ consists of a finite number of singletons $\{\theta_1, \theta_2, \ldots, \theta_m\}$. If the covering number is finite, we can focus on a finite set of functions $\{g(\cdot,\theta_\ell)\}_{\ell=1}^{m}$. Those $g(\cdot,\theta)$ in the ε-neighborhood of $g(\cdot,\theta_\ell)$ will behave similarly to the center $g(\cdot,\theta_\ell)$.

If a pair of functions $l(x)$ and $u(x)$ satisfies $\|u - l\|_{Q,r} \leq \varepsilon$ and $l(x) \leq u(x)$, we call (l, u) an ε-$L_r(Q)$ **bracket**. The **bracketing number**, denoted as $N_{[]}(\varepsilon, L_r(Q))$, is the smallest number of brackets such that for each $g(\cdot,\theta) \in \mathcal{G}$ there exists a bracket (l_ℓ, u_ℓ) satisfying $l_\ell(x) \leq g(x,\theta) \leq u_\ell(x)$. The upper bracket u_ℓ or the lower bracket l_ℓ does not have to belong to $\mathcal{G}$.

Example 1.3. The covering number and the bracketing number on the functional space are somewhat abstract. Let us use a trivial example of the unit real interval $[0, 1]$ to demonstrate the spirit behind it. If we set $\varepsilon = 0.1$, we can pick up 5 points $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ so that the 0.1-neighborhoods of these points covers the entire unit interval. The covering number in this case is 5. Regarding bracketing, $[0, 0.1], [0.1, 0.2], \ldots, , [0.9, 1]$ are 0.1-brakets and they capture all points in the unit interval. The 0.1-bracketing number is 10.

Bracketing number is more restrictive than covering number. For any fixed r, we consider a 2ε-bracket $[l, u]$. The middle line $(l + u)/2$ decomposes the 2ε-bracket into two ε-brackets $[l, \frac{l+u}{2}]$ and $[\frac{l+u}{2}, u]$ and thus $l, u \in \mathcal{N}_{\varepsilon,r,Q}\left(\frac{l+u}{2}\right)$. For any g in the bracket $[l, u]$, it must fall into $\mathcal{N}_{\varepsilon,r,Q}\left(\frac{l+u}{2}\right)$. As a result, $N_r(\varepsilon, Q) \leq N_{[]}(2\varepsilon, L_r(Q))$.

1.3 Uniform Law of Large Numbers

Definition 1.1. The sample average $\bar{g}_n(\theta)$ satisfies the **uniform law of large numbers (ULLN)** over Θ if

$$\sup_{\theta\in\Theta} |\bar{g}_n(\theta) - E[g(X_i,\theta)]| \xrightarrow{p} 0.$$

Below is a set of sufficient conditions for ULLN.

Theorem 1.1 (ULLN)**.** *If*
1. X_i *are i.i.d.*
2. $E[G(X)] < \infty$.
3. One of the following holds:
(i) $N_{[\,]}(\varepsilon, L_1(Q)) < \infty$ *for all* $\varepsilon > 0$.
(ii) $N_1(\varepsilon, Q) < \infty$ *for all* $\varepsilon > 0$.
(iii) Θ *is compact and* $P\{\omega \in \Omega : g(X(\omega),\theta)$ *is continuous in* $\theta\} = 1$*at every* $\theta \in \Theta$.
then ULLN holds.

Condition 1 is for simplicity. Condition 2 allows us to invoke dominated convergence theorem. Condition 3(i) and (ii) are about finite covering number and bracketing number, respectively. See Newey and MacFadden (1994, Handbook of Econometrics, Chapter 36)'s Lemma 2.4 for Condition 3(iii).

Example 1.4. We verify that of ECDF satisfies Theorem 1.1. For $g(x,\theta) = \mathbb{I}(x \le \theta)$, an envelop function is $G(x) = 1$, which is integrable as $E[G(x)] = 1$ for all Q. Next, we construct an ε-L_1 bracket system. To be specific, consider $\varepsilon = 1/8$. We can pick up the 0/8-th quantile $q_0 = -\infty$, 1/8-quantile q_1, 2/8-th quantile q_2 and so on up to the 7/8-th quantile q_7, and 8/8-th quantile $q_8 = \infty$, which form 8 brackets $[h_{\ell-1}, h_\ell] = [\mathbb{I}(x \le q_{\ell-1}), \mathbb{I}(x \le q_\ell)]$, for $\ell = 1, \ldots 8$. ($h_0 = 0$ and $h_8 = 1$ are outside of $\{\mathbb{I}(x \le \theta) : \theta \in \mathbb{R}\}$). It is easy to check that each $[h_{\ell-1}, h_\ell]$ is a $1/8$-L_1 bracket, and $g(x,\theta) = \mathbb{I}(x \le \theta)$ must fall into one of the brackets. In general, for any $\varepsilon > 0$, we can construct such a bracket system based on the quantiles. The bracketing number is $\lfloor 1/\varepsilon \rfloor$.

Since the bracketing system in Example 1.4 is independent of the underlying probability distribution Q, we have the well-known Glivenko-Cantelli Theorem.

Corollary 1.1 (Glivenko-Cantelli Theorem)**.** *If* X_i *is iid, then* $\sup_{\theta\in\mathbb{R}} |F_n(\theta) - F(\theta)| \xrightarrow{p} 0$.

1.4 Functional Central Limit Theorem

In earlier lectures, CLT is pointwise about $\nu_n(\theta)$ for a fixed θ. In this lecture, we consider $\nu_n(\theta)$ as a process indexed by θ as an entity.

For two non-random processes $\nu_1(\theta)$ and $\nu_2(\theta)$, we define a **uniform metric** $\rho(\nu_1,\nu_2) = \sup_{\theta\in\Theta} |\nu_1(\theta) - \nu_2(\theta)|$ as a measure of the distance. Let V be a class of processes $\nu : \Theta \to \mathbb{R}$.

Definition 1.2. Convergence in distribution: $\nu_n \xrightarrow{d} \nu$ if for every bounded, continuous $f : V \to \mathbb{R}$, we have $E[f(\nu_n)] \to E[f(\nu)]$, where continuity is defined with respect to the uniform metric.

Example 1.5. Consider the ECDF example

$$\nu_n = \sqrt{n}(F_n(\theta) - F(\theta))$$

[diagram] one realization. A candidate f is

$$f(\nu_n(\theta)) = \int_{[0,1]} (|\nu_n(\theta)| \wedge M)\, \mathrm{d}\theta$$

where the constant M and the integration domain $[0,1]$ set the upper bound $f(\nu_n(\theta)) \leq M$, and the integration is with respect to the Lebesgue measure. Another function can be

$$f(\nu_n(\theta)) = \int (\nu_n(\theta) \wedge M)^2\, \mathrm{d}\Phi(\theta)$$

where $\Phi(\theta)$ is the CDF of $N(0,1)$.

Definition 1.3. A random function $S_n(\theta)$ is **stochastic equicontinuous** if for each pair of $\varepsilon, \eta > 0$, we have $\delta > 0$ such that

$$\limsup_{n\to\infty} \Pr\left(\sup_{\theta\in\Theta} \sup_{\theta'\in N_\delta(\theta)} \left\|S_n(\theta) - S_n(\theta')\right\| > \eta\right) \leq \varepsilon$$

where $N_\delta(\theta) = \{\theta' \in \Theta : \|\theta - \theta'\| \leq \delta\}$ is a neighbor of $\theta \in \Theta$.

[diagram]
[diagram]

It allows paths with jumps, as long as the probability associated with the jumps are small enough. It's an asymptotic and probabilistic generalization of uniform continuity.

Theorem 1.2 (FCLT). *$\nu_n \xrightarrow{d} \nu$ over $\Theta = U_{j=1}^J \Theta_j$ if and only if*

1. *$(\nu_n(\theta_1), \nu_n(\theta_2), ..., \nu_n(\theta_m)) \xrightarrow{d} (\nu(\theta_1), \nu(\theta_2), ..., \nu(\theta_m))$ for every finite set $\theta_1, \theta_2, ..., \theta_m \in \Theta$*
2. *ν_n is stochastic equicontinuous over each $\Theta_1, \Theta_2, ..., \Theta_J$.*

Stochastic equicontinuity if difficult to verify, a sufficient condition is related to bracket integral

$$J_{[\,]}(\delta, L_2(Q)) = \int_0^\delta \sqrt{\log N_{[\,]}(\varepsilon, L_2(Q))}\mathrm{d}\varepsilon$$

Example 1.6. If $N_{[\,]}(\varepsilon, L_2(Q)) = \varepsilon^{-\rho}$ for some $\rho > 0$, then $\sqrt{\log N_{[\,]}(\varepsilon, L_2(Q))} = \rho^{1/2}\left(\log \frac{1}{\varepsilon}\right)^{1/2}$. As $\varepsilon \to 0$, obviously $\log \frac{1}{\varepsilon} \to \infty$. The increase is slow in that $\int_0^1 \sqrt{\log \frac{1}{\varepsilon}}\mathrm{d}\varepsilon = \frac{\sqrt{\pi}}{2}$.

Theorem 1.3. *$\nu_n(\theta)$ is stochastic equicontinuous if $J_{[\,]}(1, L_2(Q)) < \infty$, or $J_2(1) = \int_0^1 \sqrt{\log N_2(\varepsilon, Q)}\mathrm{d}\varepsilon < \infty$ and $E\left[G^2(x)\right] < \infty$.*

1.5 Donsker's Theorem

We characterize the distribution of the empirical CDF. Let

$$g(X, \theta) = \mathbb{I}(X \leq \theta) = \begin{cases} 1, & \textit{with probability } F(\theta) \\ 0, & \textit{with probability } 1 - F(\theta) \end{cases}$$

and the demeaned version is $U_i(\theta) = \mathbb{I}(X \leq \theta) - F(\theta)$. The variance of $U_i(\theta)$ is

$$var(U_i(\theta)) = (1 - F(\theta))\, F(\theta)$$

and the covariance is

$$Cov\left(U_i\left(\theta_1\right),U_i\left(\theta_2\right)\right)=E\left[\mathbb{I}\left(X\leq\theta_1\right)\mathbb{I}\left(X\leq\theta_2\right)\right]-F\left(\theta_1\right)F\left(\theta_2\right)=F\left(\theta_1\wedge\theta_2\right)-F\left(\theta_1\right)F\left(\theta_2\right).$$

Since $U_i\left(\theta\right)$ is i.i.d,

$$\left(\begin{array}{c}\nu_n\left(\theta_1\right)\\ \nu_n\left(\theta_2\right)\end{array}\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\begin{array}{c}U_i\left(\theta_1\right)\\ U_i\left(\theta_2\right)\end{array}\right)\xrightarrow{d}N\left(\left(\begin{array}{c}0\\0\end{array}\right),\left(\begin{array}{cc}F\left(\theta_1\right)\left(1-F\left(\theta_1\right)\right) & F\left(\theta_1\wedge\theta_2\right)-F\left(\theta_1\right)F\left(\theta_2\right)\\ F\left(\theta_1\wedge\theta_2\right)-F\left(\theta_1\right)F\left(\theta_2\right) & F\left(\theta_2\right)\left(1-F\left(\theta_2\right)\right)\end{array}\right)\right)$$

The joint normal distribution can be extended to for any finite $\theta_1,\theta_2,...,\theta_m$.

In the special case of $X\sim Uniform\left(0,1\right)$, the above limit distribution becomes

$$N\left(\left(\begin{array}{c}0\\0\end{array}\right),\left(\begin{array}{cc}\theta_1\left(1-\theta_1\right) & \theta_1\wedge\theta_2-\theta_1\theta_2\\ \theta_1\wedge\theta_2-\theta_1\theta_2 & \theta_2\left(1-\theta_2\right)\end{array}\right)\right)$$

This special process is called **Brownian bridge**, denoted as

$$\widetilde{B}\left(r\right)=B\left(r\right)-rB\left(1\right)$$

where $B\left(r\right)$ for $r\in[0,1]$ is a Brownian motion on the unit interval.