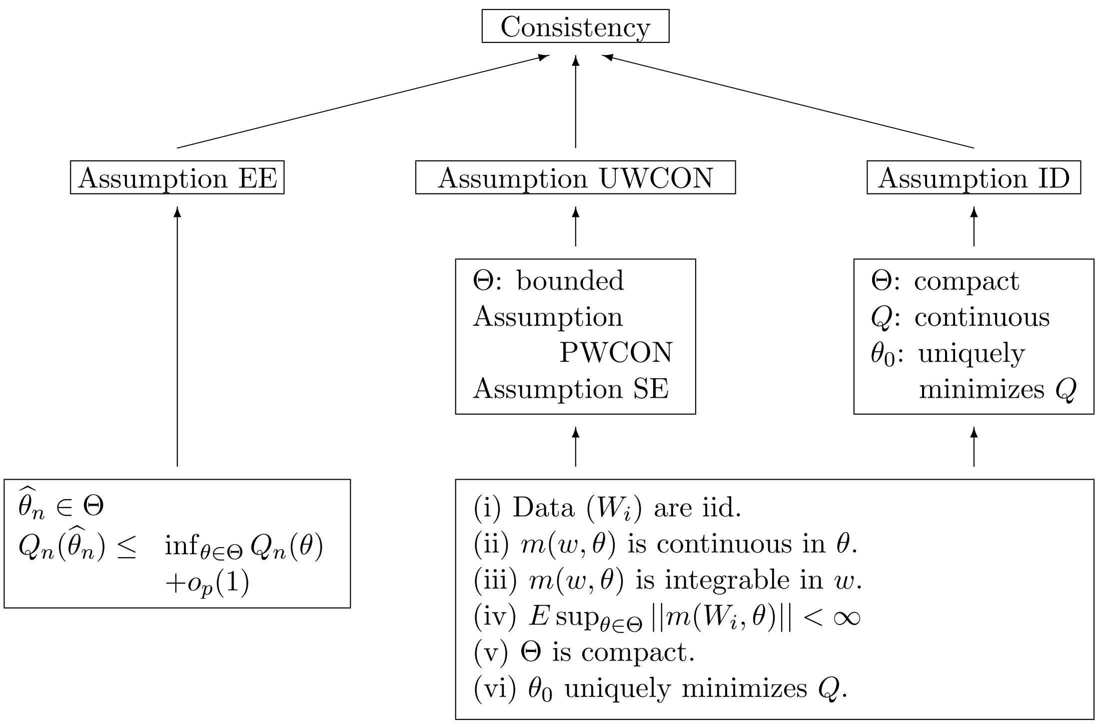
TA section 1

April 3rd, 2009 by Yang

1. Big Picture

We want to find the conditions that guarantee the consistency of extremum estimators.



2. MLE example

What is Maximum Likelihood Estimation? We want to maximize the joint likelihood of the data by chooing apporpriate parameters.

$$\text{Joint Likelihood} = \text{Joint Probability Density (Mass)}$$

Assuming iid data,

$$\max_\theta L(W;\theta) = \prod_{i=1}^{n} f(W_i;\theta)$$

or equivalently,

$$\max_\theta l(W;\theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(W_i;\theta)$$

To be consistent with the other estimation methods, we alternatively,

$$\min_{\theta} -l(W;\theta) = -\frac{1}{n}\sum_{i=1}^{n} \log f(W_i;\theta)$$

Here we set $Q_n(\theta) := -l(W;\theta)$ and $m(W_i,\theta) := -\log f(W_i;\theta)$, then

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(W_i,\theta)$$

and by weak LLN, it converges to

$$Q(\theta) = Em(W_i,\theta)$$

in probability. (PWCON under the assumption that W_i are iid, and $E||\log f(W_i;\theta)|| < \infty$)

Let us investigate other primitive assumptions.

- Is $m(w,\theta)$ continuous in θ? Yes, as long as f is continuous in θ.
- Is $m(w,\theta)$ integrable in w? Yes, as long as f is integrable in w.
- Is Θ compact? Yes, we can always find such a Θ.
- Does θ_0 uniquely minimizes Q? Yes, we proved it, under some assumption.

For example, let

$$f(W_i;\theta) = \frac{1}{\sqrt{2^d\pi^d|\Sigma|}} \exp\left(-\frac{1}{2}(W_i-\mu)'\Sigma^{-1}(W_i-\mu)\right)$$

then

$$m(W_i,\theta) = \frac{d}{2}\log(2\pi) + \frac{1}{2}\log|\Sigma| + \frac{1}{2}(W_i-\mu)'\Sigma^{-1}(W_i-\mu)$$

Check the above assumptions.

3. 2SLS Estimation example

Sometimes we may not have $EX_iU_i = 0$, but $EZ_iU_i = 0$, where $Y_i = X_i'\theta + U_i$. So we want to find θ that satisfies

$$EZ_i(Y_i - X_i'\theta) = 0$$

or by using the sample analogue,

$$\frac{1}{n}\sum_{i=1}^{n} Z_i(Y_i - X_i'\theta) = 0$$

When $dim(Z_i) = dim(\theta)$, this is easy to solve, since there are as many variables as equations. If $dim(Z_i) > dim(\theta)$, we want to minimize the difference in some way. One way to do this is minimzing sum of the squares of the difference with some weights.

For instance, if we use the weight $(Z'Z/n)^{-1/2}$,

$$\min_\theta \frac{1}{2}\left\|\left(\frac{1}{n}\sum_{i=1}^n Z_i Z_i'\right)^{-\frac{1}{2}}\frac{1}{n}\sum_{i=1}^n Z_i(Y_i - X_i'\theta)\right\|^2$$

which is the same as

$$\min_\theta \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^n (Y_i - X_i'\theta)Z_i'\right)\left(\frac{1}{n}\sum_{i=1}^n Z_i Z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n Z_i(Y_i - X_i'\theta)\right)$$

Using matrix notation and multiplying by $2n$,

$$\min_\theta (Y - X\theta)'Z(Z'Z)^{-1}Z'(Y - X\theta)$$

The FOC is

$$-2X'Z(Z'Z)^{-1}Z'(Y - X\theta) = 0$$

and thus the solution is

$$\widehat{\theta} = (X'P_Z X)^{-1}X'P_Z Y$$

which is called 2 Stage Least Squares estimator.

This is a special case of Generalized Method of Moments estimation. Let $A_n = (Z'Z/n)^{-1/2}$, and $g(W_i, \theta) = Z_i(Y_i - X_i'\theta)$, then

$$Q_n(\theta) := \frac{1}{2}\left\|A_n \frac{1}{n}\sum_{i=1}^n g(W_i, \theta)\right\|^2$$

Indeed, let $m(W_i, \theta) := g(W_i, \theta)$, then

$$Q_n(\theta) = \frac{1}{2}\left\|A_n \frac{1}{n}\sum_{i=1}^n m(W_i, \theta)\right\|^2$$

and by weak LLN, it converges to

$$Q(\theta) = \frac{1}{2}\left\|AEm(W_i, \theta)\right\|^2$$

where $A_n \xrightarrow{p} A$. Using the explicit form of functions,

$$Q(\theta) = \frac{1}{2}\left\|(EZ_i Z_i')^{-1/2}EZ_i(Y_i - X_i'\theta)\right\|^2$$

So PWCON requires iid Y_i, X_i, and Z_i, $E||Z_i Z_i'|| < \infty$ and $E||Z_i(Y_i - X_i'\theta)|| < \infty$.

Does θ_0 uniquely minimize Q? The answer would be yes, if we assume $EX_i Z_i'(EZ_i Z_i')^{-1}EZ_i X_i'$ is invertible. We will come back to this issue later in the lecture. Check also the other primitive assumptions.

4. Appendix

(1) The norm operator $\|\cdot\|$

With a $k \times 1$ vector x, $||x|| = \sqrt{x_1^2 + \cdots + x_k^2}$. $E||m(W_i, \theta)|| < \infty$ is true if the second moment is finite.

(2) Weak convergence

Weak convergence refers to convergence in probability, in comparison to almost sure convergence which is often called as strong convergence. Weak law of large numbers is a version of LLN that gives a result of weak convergence. Weak LLN holds if the data are iid, and the second moment of the summand is finite.

(3) O_p and o_p

In math, we use the notation O(big Oh) and o(little oh). $O(x^2)$ means a function at most of order x^2. e.g. $5x^2 + 7x - 3$ is $O(x^2)$. More formally, $O(x^2)$ is a function that is finite when we divide it by x^2 and let $x \to \infty$. $o(x^2)$ means a function of order less than x^2. More formally, $o(x^2)$ is a function that converges to 0 when we divide it by x^2 and let $x \to \infty$.

In probability space, we can use similar notations to denote a sequence of random variables. O_p(big Oh pi) is a sequence that is bounded in probability, and o_p(little oh pi) is one that converges to 0 in probability. e.g. $X_n = O_p(n^3)$, then X_n diverges as the same rate as n^3. e.g. $X_n = o_p(1)$, then X_n converges to 0 in probability as $n \to \infty$.

$$O_p(1) + O_p(1) = O_p(1)$$
$$O_p(1) + o_p(1) = O_p(1)$$
$$o_p(1) + o_p(1) = o_p(1)$$
$$O_p(1)O_p(1) = O_p(1)$$
$$O_p(1)o_p(1) = o_p(1)$$
$$o_p(1)o_p(1) = o_p(1)$$

There are more, e.g. $O_p(n^2) + o_p(n) = O_p(n^2)$, $o_p(n^2) + O_p(n) = o_p(n^2)$, $O_p(n^2)o_p(n) = o_p(n^3)$, and so on.

(4) sup and inf

These are almost the same as max and min, except that they always exist while max and min may not. Let $A \subset \mathbb{R}$ be some set. $\sup A$ is the smallest number no less than any element in A. $\inf A$ is the biggest number no greater than any element in A.

$\limsup_{n\to\infty} A_n$ is defined as $\lim_{n\to\infty} \sup_{k\geq n} A_k$. This always exists. $\liminf$ is defined similary. $\lim A_n$ exists only when $\limsup A_n = \liminf A_n$, in which case

$$\lim_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$$