GMM with nonsmooth moments

Class Notes

Manuel Arellano

September 25, 2009

1 Introduction

Standard GMM distribution theory requires differentiability of the moment functions. Differentiability is used to establish asymptotic normality by expanding the sample moments evaluated at the estimator around true values. However, there are important estimation problems where the sample moments are non-differentiable or even discontinuous. These include quantile methods and simulated moment estimators.

Typically population moments will be differentiable even if sample moments are not. It turns out that in such situations we can mimic the traditional approach under certain regularity conditions, which effectively allow us to interchange the order of expectation and differentiation. That is, we may be able to approximate the scaled estimation error by a linear combination of sample moments at the truth, with weights given by derivatives of expectations as opposed to expectations of derivatives.

Here we discuss a set of regularity conditions under which asymptotic normality can be established for consistent estimators that approximately solve nondifferentiable moments equations. This leads us to describe the general form of asymptotic variances and how to obtain them. We conclude with the statement of a formal asymptotic normality theorem taken from Newey and McFadden (1994). As a simple illustration we discuss the estimation of unconditional quantiles throughout.

2 Unconditional quantiles

Let $F(r) = \Pr(Y \leq r)$. For $\tau \in (0,1)$, the τth population quantile of Y is defined to be

$$q_\tau \equiv F^{-1}(\tau) = \inf\{r : F(r) \geq \tau\}.$$

$F^{-1}(\tau)$ is a generalized inverse function. It is a left-continuous function with range equal to the support of F and hence often unbounded

Let us define the "check" function (or asymmetric absolute loss function). For $\tau \in (0,1)$

$$\rho_\tau(u) = [\tau \mathbf{1}(u \geq 0) + (1-\tau)\mathbf{1}(u < 0)] \times |u| = [\tau - \mathbf{1}(u < 0)]\, u$$

Note that $\rho_\tau(u)$ is a continuous piecewise linear function, but nondifferentiable at $u = 0$. We should think of u as an individual error $u = y - r$ and $\rho_\tau(u)$ as the loss associated with u.

Using $\rho_\tau(u)$ as a specification of loss, it is well known that q_τ minimizes expected loss:

$$s_0(r) \equiv E[\rho_\tau(Y - r)] = \tau \int_r^\infty (y-r)\, dF(y) - (1-\tau)\int_{-\infty}^r (y-r)\, dF(y).$$

Any element of $\{r : F(r) = \tau\}$ minimizes expected loss. If the solution is unique, it coincides with q_τ as defined above. If not, we have an interval of τth quantiles and the smallest element is chosen so that the quantile function is left-continuous (by convention).

Given a random sample $\{Y_1, ..., Y_n\}$ we obtain sample quantiles replacing F by the empirical cdf:

$$F_n(r) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(Y_i \leq r).$$

That is, we choose $\widehat{q}_\tau = F_n^{-1}(\tau) \equiv \inf\{r : F_n(r) \geq \tau\}$, which minimizes

$$s_n(r) = \int \rho_\tau(y - r)\, dF_n(y) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - r).$$

An important advantage of expressing the calculation of sample quantiles as an optimization problem, as opposed to a problem of ordering the observations, is computational (specially in the regression context). The optimization perspective is also useful for studying statistical properties.

The sample objective function $s_n(r)$ is continuous but not differentiable for all r. Moreover, the gradient or moment condition

$$b_n(r) = \frac{1}{n}\sum_{i=1}^{n} [\mathbf{1}(Y_i \leq r) - \tau]$$

is not continuous in r. Note that if each Y_i is distinct, so that we can reorder the observations to satisfy $Y_1 < Y_2 < ... < Y_n$, for all τ we have

$$|b_n(\widehat{q}_\tau)| \equiv |F_n(\widehat{q}_\tau) - \tau| \leq \frac{1}{n}.$$

We can now illustrate the point that, despite lack of smoothness in $s_n(r)$ or $b_n(r)$, smoothness of the distribution of the data can smooth their population counterparts. Suppose that F is differentiable at q_τ with positive derivative $f(q_\tau)$, then $s_0(r)$ is twice continuously differentiable with derivatives:

$$\frac{d}{dr}E[\rho_\tau(Y - r)] = -\tau[1 - F(r)] + (1 - \tau)F(r) = F(r) - \tau \equiv E[\mathbf{1}(Y \leq r) - \tau]$$

$$\frac{d^2}{dr^2}E[\rho_\tau(Y - r)] = f(r).$$

3 Consistency

The basic requirements for consistency of an extremum estimator are that the limiting objective function is uniquely maximized at the truth (identification), boundedness of the parameter set, continuity of the objective function, and uniform convergence. Here we state a slightly different version to Amemiya (1985)'s consistency theorem, which replaces the requirement of continuity of the objective function by continuity of the limiting objective function. In this way the theorem covers, for example, GMM objective functions based on sample moments that are not continuous. The theorem is taken from Newey and McFadden (1994, Theorem 2.1).

Consistency Theorem Suppose that $\widehat{\theta}$ maximizes the objective function $S_n(\theta)$ in the parameter space Θ. Assume the following:

(a) Θ is a compact.

(b) $S_n(\theta)$ converges uniformly in probability to $S_0(\theta)$.

(c) $S_0(\theta)$ is continuous.

(d) $S_0(\theta)$ is uniquely maximized at θ_0.

Then $\widehat{\theta} \xrightarrow{p} \theta_0$. (Proof: See Newey and McFadden, 1994, p. 2121–2).

Example Consistency of sample quantiles follows from this theorem under fairly general assumptions. The quantile sample objective function $s_n(r)$ is continuous and convex in r. Suppose that F is such that $s_0(r)$ is uniquely maximized at q_τ. By the law of large numbers $s_n(r)$ converges pointwise to $s_0(r)$. Then use the fact that pointwise convergence of convex functions implies uniform convergence on compact sets.

Alternatively, one can argue that the requirement of uniform convergence of the criterion function:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - r) = \int \rho_\tau(y - r)\, dF(y) \text{ uniformly in } r,$$

is guaranteed by the Glivenko-Cantelli theorem, which establishes that the empirical cdf $F_n(r)$ from an iid sequence with cdf $F(r)$, converges a.s. uniformly to $F(r)$.[1]

4 Asymptotic normality

The asymptotic normality of sample quantiles cannot be established in the standard way because of the nondifferentiability of the objective function. However, it has long been known that under suitable conditions sample quantiles are asymptotically normal and there are direct approaches to establish the result. See for example the proofs in Cox and Hinckley (1974, p. 468) and Amemiya (1985, p. 148–150). But here quantiles are only used to illustrate the applicability of a general approach.

We seek a methodology of proof that resembles as closely as possible the familiar approach for differentiable problems. It turns out that as long as the limiting objective function is differentiable such approach is possible if a reminder term in the approximating argument is sufficiently small. We first discuss the argument in a just-identified GMM problem for simplicity.

[1] See Amemiya (1985, p. 150) for a proof of consistency of the median, and Koenker (2005, p. 117–119) for conditional and unconditional quantiles. A simple proof of the Glivenko-Cantelli theorem is given in van der Vaart (1998, p. 266).

Consider a population quantity $\theta_0 \in \mathbb{R}^k$ that satisfies a set of k moments or first-order conditions

$$\overline{b}(\theta) \equiv E\psi(W,\theta)$$

so that $\overline{b}(\theta_0) = 0$. Also consider a consistent estimator $\widehat{\theta}$ that approximately satisfies the sample moment conditions, in the sense that

$$\lim_{n\to\infty} \sqrt{n} b_n\left(\widehat{\theta}\right) \overset{a.s.}{\to} 0, \tag{1}$$

where

$$b_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\psi(W_i,\theta).$$

We are interested in situations where $\psi(W_i,\theta)$ is not differentiable in θ and possibly not continuous (such as $\psi(W,\theta) = 1(W \leq \theta) - \tau$). In these situations it is not possible to establish asymptotic normality of $\widehat{\theta}$ by expanding $\sqrt{n}b_n\left(\widehat{\theta}\right)$ around θ_0, as is usual in differentiable moment problems. Provided $\overline{b}(\theta)$ is differentiable in θ, consider instead a mean value expansion of $\overline{b}(\theta_0)$ around $\widehat{\theta}$:

$$0 = \sqrt{n}\overline{b}(\theta_0) = \sqrt{n}\overline{b}\left(\widehat{\theta}\right) - \frac{\partial \overline{b}\left(\widetilde{\theta}\right)}{\partial \theta'}\sqrt{n}\left(\widehat{\theta} - \theta_0\right)$$

where $\widetilde{\theta}$ is such that $\left\|\widetilde{\theta} - \theta_0\right\| \leq \left\|\widehat{\theta} - \theta_0\right\|$ and takes different values for each column of $\partial\overline{b}\left(\widetilde{\theta}\right)/\partial\theta'$. With iid observations and continuity of $\partial\overline{b}(\theta)/\partial\theta'$ at θ_0:

$$\operatorname*{plim}_{n\to\infty} \frac{\partial \overline{b}\left(\widetilde{\theta}\right)}{\partial \theta'} = \frac{\partial \overline{b}(\theta_0)}{\partial \theta'} \equiv D_0.$$

Thus, provided D_0 is nonsingular

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) = \left[D_0^{-1} + o_p(1)\right]\sqrt{n}\overline{b}\left(\widehat{\theta}\right)$$

Next, let us define the empirical process at $\psi(W_i,\theta)$:

$$v_n(\theta) = \sqrt{n}\left[b_n(\theta) - \overline{b}(\theta)\right]$$

Note that for fixed $\theta \neq \theta_0$, $b_n(\theta) - \overline{b}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[\psi(W_i,\theta) - \overline{b}(\theta)\right]$ is a sample average of zero-mean iid variables so that $v_n(\theta)$ is $O_p(1)$ and asymptotically normal.

Using this notation we can write

$$\begin{aligned} -\sqrt{n}\overline{b}\left(\widehat{\theta}\right) &= \left[\sqrt{n}b_n\left(\widehat{\theta}\right) - \sqrt{n}\overline{b}\left(\widehat{\theta}\right)\right] - \sqrt{n}b_n\left(\widehat{\theta}\right) \\ &= \left[v_n\left(\widehat{\theta}\right) - v_n(\theta_0)\right] + v_n(\theta_0) - \sqrt{n}b_n\left(\widehat{\theta}\right) \end{aligned}$$

or in view of (1) and $v_n(\theta_0) \equiv \sqrt{n}b_n(\theta_0)$, also

$$-\sqrt{n}\overline{b}\left(\widehat{\theta}\right) = \sqrt{n}b_n(\theta_0) + \left[v_n\left(\widehat{\theta}\right) - v_n(\theta_0)\right] + o_p(1).$$

Note that under standard conditions,

$$\sqrt{n} b_n (\theta_0) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi (W_i, \theta_0) \xrightarrow{d} \mathcal{N} (0, V_0)$$

where

$$V_0 = E \left[\psi (W_i, \theta_0) \psi (W_i, \theta_0)' \right].$$

The critical assumption is

$$v_n \left(\widehat{\theta} \right) - v_n (\theta_0) \equiv \sqrt{n} \left[b_n \left(\widehat{\theta} \right) - b_n (\theta_0) - \overline{b} \left(\widehat{\theta} \right) \right] = o_p (1). \tag{2}$$

An empirical process that satisfies this type of condition is said to be stochastically equicontinuous. Note that stochastic equicontinuity requires that $\sqrt{n} \left[b_n (\theta_0) - \overline{b} (\theta_0) \right]$ and $\sqrt{n} \left[b_n \left(\widehat{\theta} \right) - \overline{b} \left(\widehat{\theta} \right) \right]$ are asymptotically equivalent.

Thus, under (2) we have

$$\sqrt{n} \left(\widehat{\theta} - \theta_0 \right) = -D_0^{-1} \sqrt{n} b_n (\theta_0) + o_p (1) \xrightarrow{d} \mathcal{N} \left(0, D_0^{-1} V_0 D_0^{-1} \right).$$

The previous discussion follows Andrews (1994, p. 2255–2258)' heuristic description of the problem. A general theorem for overidentified GMM due to Newey and McFadden (1994) is stated below. In the proofs of these results the most difficult condition to check is the stochastic equicontinuity assumption, which asserts that a scaled difference in differences expression between sample and expected moments evaluated at the estimator and the truth is of small order. General results in the literature of empirical processes tend to rely on a regularity condition of stochastic equicontinuity, but finding more primitive conditions that are applicable to specific problems is often difficult. An early contribution in this line of reasoning is Huber (1967). In econometrics, Powell (1984, 1986) applied Huber's conditions to censored regression quantiles. See also Amemiya (1985)'s argument for LAD (p. 154).

Asymptotic Normality Theorem Suppose that $b_n \left(\widehat{\theta} \right)' A_n b_n \left(\widehat{\theta} \right) \leq \inf_{\theta \in \Theta} b_n (\theta)' A_n b_n (\theta) + o_p \left(n^{-1} \right)$, $\widehat{\theta} \xrightarrow{p} \theta_0$, and $A_n \xrightarrow{p} A_0 \geq 0$, where there is $\overline{b} (\theta)$ such that

1. $\overline{b} (\theta_0) = 0$;
2. $\overline{b} (\theta)$ is differentiable at θ_0 with derivative D_0 such that $D_0' A_0 D_0$ is nonsingular:
3. θ_0 is an interior point of Θ.
4. $\sqrt{n} b_n (\theta_0) \xrightarrow{d} \mathcal{N} (0, V_0)$;

5. Stochastic equicontinuity assumption: For any $\delta_n \to 0$,

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \frac{\sqrt{n}\left\|b_n(\theta)-\overline{b}(\theta)-b_n(\theta_0)\right\|}{1+\sqrt{n}\|\theta-\theta_0\|} \xrightarrow{p} 0.$$

Then,

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right) \xrightarrow{d} \mathcal{N}\left[0,\left(D_0' A_0 D_0\right)^{-1} D_0' A_0 V_0 A_0 D_0 \left(D_0' A_0 D_0\right)^{-1}\right].$$

(Proof: See Newey and McFadden, 1994, Theorem 7.2, p. 2186).

In this theorem $b_n(\theta)$ is allowed to be discontinuous. The theorem is also similar to Huber (1967) and Pakes and Pollard (1989) in a case where the number of moments and parameters coincide. Newey and McFadden also provide a more general theorem for extremum estimators (their Theorem 7.1), which covers the previous one as a special case. Although not made explicit in the theorem, the function $\overline{b}(\theta)$ should be thought of as the limit of $b_n(\theta)$.

Newey and McFadden's form of the stochastic equicontinuity assumption (condition 5) is similar to Huber (1967)'s and sightly more general than the alternative form consisting of condition 5 without the denominator term:

$$\sup_{\|\theta-\theta_0\|\leq\delta_n} \sqrt{n}\left\|b_n(\theta)-\overline{b}(\theta)-b_n(\theta_0)\right\| \xrightarrow{p} 0. \tag{3}$$

Andrews (1994) discusses more primitive conditions that lead to this result. Andrews (1994) also discusses two equivalent definitions of stochastic equicontinuity (see appendix), which make explicit the connection between (2) and (3).

Example The asymptotic normality result for unconditional quantiles given below, follows directly from the previous discussion, once the stochastic equicontinuity condition is checked for the empirical process

$$v_n(\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\mathbf{1}(Y_i \leq \theta) - F(\theta)\right].$$

Letting

$$r_n(\theta,\theta_0) = \sqrt{n}\left[b_n(\theta)-\overline{b}(\theta)-b_n(\theta_0)\right] = v_n(\theta) - v_n(\theta_0),$$

note that for any θ

$$E\left[r_n(\theta,\theta_0)\right] = 0$$

$$Var\left[r_n(\theta,\theta_0)\right] = F(\theta)\left[1-F(\theta)\right] + F(\theta_0)\left[1-F(\theta_0)\right] - 2F(\min(\theta,\theta_0))\left[1-F(\max(\theta,\theta_0))\right],$$

so that $Var\left[r_n(\theta,\theta_0)\right]\to 0$ as $\delta_n\to 0$. Therefore, pointwise convergence in probability for $\|\theta-\theta_0\|\leq\delta_n$ is readily established: $r_n(\theta,\theta_0)\xrightarrow{p}0$.

The function $\psi(Y-\theta)=\mathbf{1}(Y-\theta\leq 0)-\tau$ belongs to the type I class of functions discussed in Andrews (1994). In such a case stochastic equicontinuity is trivially verified given that $\psi(.)$ is bounded. Andrews (p. 2273) provides formal sufficient conditions for stochastic equicontinuity of first-order conditions of M-estimators, of which unconditional quantiles is a simple special case.

Theorem: Asymptotic normality of unconditional sample quantiles Fix $0<\tau<1$. If F is differentiable at q_τ with positive derivative $f(q_\tau)$, then

$$\sqrt{n}\left(\widehat{q}_\tau-q_\tau\right)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbf{1}(Y_i\leq q_\tau)-\tau}{f(q_\tau)}+o_p(1).$$

Consequently,

$$\sqrt{n}\left(\widehat{q}_\tau-q_\tau\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\tau(1-\tau)}{\left[f(q_\tau)\right]^2}\right).$$

See van der Vaart, 1998, p. 307, Corollary 21.5.

In terms of the general notation, the example has $\widehat{\theta}=\widehat{q}_\tau$, $\theta_0=q_\tau$, $\psi(W_i,\theta_0)=\mathbf{1}(Y_i\leq q_\tau)-\tau$,

$$b_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{1}(Y_i\leq\theta)-\tau\right],\quad \overline{b}(\theta)=F(\theta)-\tau$$

$$D_0=f(q_\tau),\quad V_0=E\left\{\left[\mathbf{1}(Y_i\leq q_\tau)-\tau\right]^2\right\}=\tau(1-\tau).$$

The term $\tau(1-\tau)$ in the numerator of the asymptotic variance tends to make $\widehat{q}_\tau$ more precise in the tails, whereas the density term in the denominator tends to make $\widehat{q}_\tau$ less precise in regions of low density. Typically the latter effect will dominate so that quantiles closer to the extremes will be estimated with less precision.

A Convergence, continuity, and equicontinuity

A.1 Pointwise and uniform convergence of a sequence of (nonrandom) functions

Suppose $\Theta \subseteq \mathbb{R}^q$ and that for each $n \in N$, g_n is a function from Θ to $\mathbb{R}$. We may think of the functions g_1, g_2, g_3,..., as forming a sequence of functions $\{g_n\}$. Of course, this is very different from a sequence of real numbers, but it is still possible to formulate some idea of 'limit' in this context. There are two main ways in which we might do this. One is through the definition of *pointwise convergence* and the other through *uniform convergence*. The following definitions describe these.

Definition 1 ***(Pointwise convergence)*** *The sequence $\{g_n\}$ converges pointwise to the function g on Θ if and only if for each $\theta \in \Theta$, $g_n(\theta) \to g(\theta)$ as $n \to \infty$; that is, given $\theta \in \Theta$, and $\varepsilon > 0$, there exists $\overline{n} = \overline{n}(\theta, \varepsilon)$ such that*

$$n > \overline{n}(\theta, \varepsilon) \Rightarrow |g_n(\theta) - g(\theta)| < \varepsilon.$$

Definition 2 ***(Uniform convergence)*** *The sequence $\{g_n\}$ converges uniformly to the function g on Θ if and only if given $\varepsilon > 0$, there exists $\overline{n} = \overline{n}(\varepsilon)$ such that*

$$n > \overline{n}(\varepsilon) \Rightarrow |g_n(\theta) - g(\theta)| < \varepsilon, \quad \text{for all } \theta \in \Theta.$$

The difference between these definitions is that, with uniform convergence, the $\overline{n}$ depends only on ε so that the same $\overline{n}$ works for every $\theta \in \Theta$, whereas in pointwise convergence, θ is given, and $\overline{n}$ can depend on θ as well as on ε. Uniform convergence is a stronger property than pointwise convergence. The former implies the latter but not the converse. The following example illustrates this point.

Suppose that we take $\Theta = [0, 1]$ and $g_n(\theta) = \theta^n$. Then it is easy to see that $\{g_n\}$ converges pointwise on $[0, 1]$ to the function $g(\theta) = \mathbf{1}(\theta = 1)$, but the convergence is not uniform because the rate of convergence is not independent of θ.

It is easy to show that the sequence $\{g_n(\theta)\}_{n=1}^{\infty}$ converges uniformly to $g(\theta)$ if and only if

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} |g_n(\theta) - g(\theta)| = 0$$

(See André Khuri (1993), *Advanced Calculus with Applications in Statistics*, Wiley, Theorem 5.3.1, p. 166).

A.2 Pointwise and uniform convergence of a sequence of random functions

Definition 3 ***(Pointwise convergence in probability)*** *The random sequence $\{g_n(\theta)\}$ converges pointwise in probability to the function $g(\theta)$ on Θ if and only if for each $\theta \in \Theta$, $|g_n(\theta) - g(\theta)| \xrightarrow{p} 0$ as $n \to \infty$, or equivalently, if and only if for each $\theta \in \Theta$,*

$$\lim_{n\to\infty} \Pr(|g_n(\theta) - g(\theta)| > \delta) = 0 \text{ for any } \delta > 0.$$

That is, given $\theta \in \Theta$ and $\varepsilon > 0$, there exists $\overline{n} = \overline{n}(\theta, \varepsilon)$ such that

$$n > \overline{n}(\theta, \varepsilon) \Rightarrow \Pr(|g_n(\theta) - g(\theta)| > \delta) < \varepsilon \text{ for any } \delta > 0.$$

Definition 4 ***(Uniform convergence in probability)*** *The random sequence $\{g_n(\theta)\}$ converges uniformly in probability to the function $g(\theta)$ on Θ if and only if $\sup_{\theta \in \Theta} |g_n(\theta) - g(\theta)| \xrightarrow{p} 0$ as $n \to \infty$, or equivalently, if and only if*

$$\lim_{n \to \infty} \Pr\left(\sup_{\theta \in \Theta} |g_n(\theta) - g(\theta)| > \delta\right) = 0 \text{ for any } \delta > 0.$$

A.3 Equicontinuity

A sequence of (nonrandom) functions is equicontinuous if all the functions are continuous and they have equal variation over a given neighborhood in the sense described below. We collect formal definitions for the purpose of comparisons with stochastic generalizations.

Equicontinuity The sequence $\{g_n(\theta)\}_{n=1}^{\infty}$ is equicontinuous at $\theta_0 \in \Theta$, if for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon, \theta_0) > 0$ such that for all n, and all $\theta \in \Theta$ such that $|\theta - \theta_0| < \delta$:

$$|\theta - \theta_0| < \delta(\epsilon, \theta_0) \Rightarrow |g_n(\theta) - g_n(\theta_0)| < \epsilon.$$

The sequence is equicontinuous if it is equicontinuous at each point of Θ.

Uniform equicontinuity The sequence $\{g_n(\theta)\}_{n=1}^{\infty}$ is uniformly equicontinuous if for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon) > 0$ such that for all n, and all $\theta_1, \theta_2 \in \Theta$ such that $|\theta_1 - \theta_2| < \delta$:

$$|\theta_1 - \theta_2| < \delta(\epsilon) \Rightarrow |g_n(\theta_1) - g_n(\theta_2)| < \epsilon.$$

That is, a sequence of functions is (uniformly) equicontinuous if they are continuous uniformly in n.

Let us compare these concepts with continuity:

Continuity $g_n(\theta)$ is a continuous function at $\theta_0 \in \Theta$, if for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon, \theta_0, n) > 0$ such that for all n, and all $\theta \in \Theta$ such that $|\theta - \theta_0| < \delta$:

$$|\theta - \theta_0| < \delta(\epsilon, \theta_0, n) \Rightarrow |g_n(\theta) - g_n(\theta_0)| < \epsilon.$$

So, for continuity, δ may depend on ϵ, θ_0 and n; for equicontinuity, δ must be independent of n; and for uniform equicontinuity, δ must be independent of both n and θ_0.

Stochastic equicontinuity The random sequence $\{g_n(\theta)\}_{n=1}^{\infty}$ is stochastically equicontinuous if and only if for every sequence of positive constants $\{\delta_n\}$ that converges to zero, we have

$$\sup_{|\theta_1-\theta_2|\leq\delta_n} |g_n(\theta_1) - g_n(\theta_2)| \xrightarrow{p} 0.$$

or equivalently, if and only if

$$\lim_{n\to\infty} \Pr\left(\sup_{|\theta_1-\theta_2|\leq\delta_n} |g_n(\theta_1) - g_n(\theta_2)| < \epsilon\right) = 1 \text{ for any } \epsilon > 0.$$

Basically, $\{g_n(\theta)\}_{n=1}^{\infty}$ is stochastically equicontinuous if $g_n(\theta)$ is continuous in θ uniformly over Θ at least with high probability and for n large.

An alternative, equivalent definition of stochastic equicontinuity is as follows:

The random sequence $\{g_n(\theta)\}_{n=1}^{\infty}$ is stochastically equicontinuous if and only if for all sequences of random elements $\left\{\widehat{\theta}_{1n}\right\}$ and $\left\{\widehat{\theta}_{2n}\right\}$ that satisfy $\left|\widehat{\theta}_{1n} - \widehat{\theta}_{2n}\right| \xrightarrow{p} 0$, we have

$$v_n\left(\widehat{\theta}_{1n}\right) - v_n\left(\widehat{\theta}_{2n}\right) \xrightarrow{p} 0.$$

See Andrews (1994, p. 2252).

References

[1] Amemiya, T. (1985): *Advanced Econometrics*, Blackwell.

[2] Andrews, D.W.K. (1994): "Empirical Process Methods in Econometrics", in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, Elsevier.

[3] Huber, P. J. (1967): "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions", in L.M. LeCam and J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley.

[4] Koenker, R. (2005), *Quantile Regression*, Cambridge University Press.

[5] Newey, W. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing", in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, Elsevier.

[6] Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57, 1027–1057.

[7] Powell, J. (1986), "Censored Regression Quantiles", *Journal of Econometrics*, 32, 143–155.

[8] van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.