Spring 1997 | John Rust
Economics 551b | 37 Hillhouse, Rm. 27

Proof of the Uniform Law of Large Numbers

This note presents a self-contained proof of the uniform strong law of large numbers (ULLN). The ULLN is useful in situations where we have sample moments of functions $h(x,\theta)$ that depends on two arguments: a random element x and a deterministic parameter θ. Suppose we observe N *IID* observations $\{\tilde{X}_1,\ldots,\tilde{X}_N\}$ from some probability distribution $F(x)$ and we fix θ at some arbitrary value in the parameter space Θ, then the ordinary strong law of large numbers (SLLN) states the following:

Strong Law of Large Numbers *If* $E\{|h(\tilde{x},\theta)|\} = \int |h(x,\theta)|F(dx) < +\infty$*, then with probability 1 we have:*

$$H_N(\theta) \equiv \frac{1}{N}\sum_{i=1}^{N} h(\tilde{x}_i,\theta) \longrightarrow E\{h(\tilde{x},\theta)\} \equiv H(\theta). \tag{1}$$

The ULLN is an extension of the SLLN that provides conditions under which $H_N(\theta)$ converges to $H(\theta)$ *uniformly* in θ, i.e. conditions under which we have:

$$\|H_N - H\| \equiv \sup_{\theta\in\Theta} |H_N(\theta) - H(\theta)| \longrightarrow 0 \quad \text{with probability 1.} \tag{2}$$

Equation (2) states that the maximum deviation between the random function H_N and the deterministic function H converges to 0: i.e. the sequence $\{H_N\}$ converges uniformly in θ (i.e. in *sup norm*) to the deterministic function H. To prove (2) it is convenient to work with the normalized functions $g(x,\theta)$ defined by

$$g(x,\theta) = h(x,\theta) - E\{h(\tilde{x},\theta)\} \tag{3}$$

Clearly $E\{g(\tilde{x},\theta)\} = 0$ for all $\theta \in \Theta$.

Uniform Strong Law of Large Numbers: *Let* $\{\tilde{x}_i\}$ *be IID random elements* $\tilde{x}_i \in X$*, where* X *is a Borel space. Let* $g(x,\theta)$ *be a measurable function of* x *for all* θ*, and a continuous function of* θ *for almost all* $x \in X$*. Suppose that* Θ *is compact, and that* $G_N(\theta) \equiv 1/N\sum_{i=1}^{N} g(x_i,\theta)$ *converges to* 0 *with probability 1 for each* $\theta \in \Theta$*. Then if* $|g(x,\theta)| < d(x)$ *for some function* d *satisfying* $E\{d(\tilde{x})\} < \infty$*, then we have* $\|G_N\| \longrightarrow 0$ *with probability 1, i.e.*

$$\sup_{\theta\in\Theta} \left|\frac{1}{N}\sum_{i=1}^{N} g(x_i,\theta)\right| \to 0 \quad \text{with probability 1.} \tag{4}$$

Proof: Define a function $\overline{g}(x,\theta,\epsilon)$ by

$$\overline{g}(x,\theta,\epsilon) \equiv \sup_{\{\theta' \,|\, |\theta'-\theta|<\epsilon\}} g(x,\theta'). \tag{5}$$

Since g is continuous in θ for almost all x it follows that for almost all x we have

$$\lim_{\epsilon\to 0} \overline{g}(x,\theta,\epsilon) = g(x,\theta). \tag{6}$$

Also, since $g(x,\theta)$ is dominated by $d(x)$ uniformly in θ, $\overline{g}(x,\theta,\epsilon)$ is also dominated by $d(x)$ and we can apply the Lebesgue dominated convergence theorem to show that

$$\lim_{\epsilon\to 0} E\{\overline{g}(\tilde{x},\theta,\epsilon)\} = E\{g(\tilde{x},\theta)\} = 0. \tag{7}$$

Similarly we can define a function $\underline{g}$ by substituting inf for sup in (5), and the result (7) will also hold for $\underline{g}$. Now consider the following inequality, which holds for all N and all sequences $\{x_i\}$ and all θ' in an ϵ-ball $B(\theta,\epsilon) = \{\theta' \big| \, |\theta'-\theta| < \epsilon\}$ about an arbitrary point $\theta \in \Theta$:

$$\frac{1}{N}\sum_{i=1}^{N} \inf_{\theta' \in B(\theta,\epsilon)} g(x_i,\theta') \le \frac{1}{N}\sum_{i=1}^{N} g(x_i,\theta') \le \frac{1}{N}\sum_{i=1}^{N} \sup_{\theta' \in B(\theta,\epsilon)} g(x_i,\theta'). \tag{8}$$

Equation (8) implies the following result

$$\sup_{\theta' \in B(\theta,\epsilon)} \Big| \frac{1}{N}\sum_{i=1}^{N} g(x_i,\theta')\Big| \le \max\left[\Big|\frac{1}{N}\sum_{i=1}^{N} \inf_{\theta' \in B(\theta,\epsilon)} g(x_i,\theta')\Big|, \Big|\frac{1}{N}\sum_{i=1}^{N} \sup_{\theta' \in B(\theta,\epsilon)} g(x_i,\theta')\Big|\right]. \tag{9}$$

Taking limits on both sides of (9) we have with probability 1:

$$\limsup_{N\to\infty} \sup_{\theta' \in B(\theta,\epsilon)} \Big| \frac{1}{N}\sum_{i=1}^{N} g(x_i,\theta')\Big| \le \max\left[\Big|E\{\underline{g}(\tilde{x},\theta,\epsilon)\}\Big|, \Big|E\{\overline{g}(\tilde{x},\theta,\epsilon)\}\Big|\right]. \tag{10}$$

Since $\underline{g}$ and $\overline{g}$ tend to 0 as $\epsilon \to 0$ by (7), given a small $\eta > 0$ choose ϵ sufficiently small that both of the terms in the *max* expression on the right hand side of (10) are less than η. Now, the collection of balls $\{B(\theta,\epsilon)\}$ form an open cover of Θ. By compactness, there is a finite subcover, $\{B(\theta_1,\epsilon),\dots,B(\theta_K,\epsilon)\}$. Since inequality (10) holds in each of these balls, we must have

$$Pr\left\{\limsup_{N\to\infty} \sup_{\theta\in\Theta} \Big|\frac{1}{N}\sum_{i=1}^{N} g(x_i,\theta)\Big| \le \eta\right\} = 1. \tag{11}$$

Taking intersections in (11) over a sequence $\{\eta_j\}$ tending to 0, it follows that

$$Pr\left\{\lim_{N\to\infty} \sup_{\theta\in\Theta} \Big|\frac{1}{N}\sum_{i=1}^{N} g(x_i,\theta)\Big| = 0\right\} = 1. \tag{12}$$

Comment: The ULLN is actually a special case of the *SLLN in Banach Spaces.* That is, the functions $g(\tilde{x}_i,\theta)$ can be regarded as *random elements* in the Banach space B of all continuous, bounded functions from Θ into R. Furthermore, each of these random elements has mean zero, i.e. the expectation of the random function $g(\tilde{x}_i,\theta)$ is the *zero function* $\mathbf{0}$, also a member of B. The SLLN in Banach spaces states the following:

SLLN for Banach Spaces *Let $\{\tilde{Z}_1,\dots,\tilde{Z}_N\}$ be IID random elements in a separable Banach space satisfying $E\{\tilde{Z}_i\} = \mathbf{0}$ (where $\mathbf{0}$ is the 0 element of B), and $E\{\|\tilde{Z}_i\|\} < +\infty$. Then we have:*

$$\frac{1}{N}\sum_{i=1}^{N} \tilde{Z}_i \longrightarrow \mathbf{0} \quad \text{with probability 1,}$$

which is equivalent to

$$\left\| \frac{1}{N} \sum_{i=1}^{N} \tilde{Z}_i \right\| \longrightarrow 0,$$

where $\|Z\|$ is the norm of the element $Z \in B$.

The ULLN emerges as a special case of the SLLN in Banach spaces by defining the Banach space B to be the space of all continuous functions from Θ to R, and the norm on B is defined as the supremum norm, i.e. the maximum the absolute value of the function as θ ranges over Θ. Then we can define random elements on B by $\tilde{Z}_i = g(\tilde{x}_i, \cdot)$ and therefore the norm of these $\tilde{Z}_i$ is given by

$$\|\tilde{Z}_i\| \equiv \sup_{\theta \in \Theta} |g(\tilde{x}_i, \theta)| .$$

It is easy to see that the random elements $\tilde{Z}_i$ satisfy the conditions of the Banach Space SLLN and therefore the sample average of these random elements converges with probability 1 to the zero element of B, $\mathbf{0}$, i.e. the 0 function. But the convergence in norm of the sample average of the $\{\tilde{Z}_i\}$ to the $\mathbf{0}$ element of B is equivalent to the uniform convergence of the sample average of the random functions $\{g(\tilde{x}_i, \theta)\}$ to the zero function, which is precisely what the ULLN states. While this more abstract approach to proving the ULLN is conceptually simpler than the direct proof given above, the mathematics involved in proving the SLLN in Banach spaces are too advanced to be covered in this handout or in Econ 551.

Comment: In general, pointwise convergence of functions does not imply uniform convergence. A classic counterexample is the (deterministic) sequence of functions g_N defined over the space $\Theta = [0, 1]$ by

$$g_N(\theta) = \begin{cases} N\theta & 0 \le \theta \le \frac{1}{N} \\ (2 - N\theta) & \frac{1}{N} < \theta \le \frac{2}{N} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

It is clear the the sequence of functions $\{g_N\}$ defined in (13) converge pointwise to the 0 function, $\mathbf{0}$, but the $\{g_N\}$ sequence can't converge uniformly to $\mathbf{0}$ since $\|g_N\| = 1$ for all N. What is going wrong here is that while each g_N is continuous, the functions are converging to a discontinous function equal to 1 at $\theta = 0$ and 0 for $\theta > 0$. Another way of saying this is that the sequence $\{g_N\}$ is not *uniformly equicontinuous.*

Definition: *A collection of functions $\mathcal{F}$ mapping Θ into R is uniformly equicontinuous if for each $\epsilon > 0$ there exists a $\delta > 0$ such that for all $f \in \mathcal{F}$ and for each θ and θ' satisfying $\|\theta - \theta'\| \le \delta$ we have:*

$$|f(\theta) - f(\theta')| < \epsilon. \tag{14}$$

The key idea of equicontinuity is that inequality (14) holds simultaneously for all $f \in \mathcal{F}$. There is a classical theorem of functional analysis, *Ascoli's Theorem* that relates uniform equicontinuity to uniform convergence:

Ascoli's Theorem: *Let $\{H_N\}$ be a sequence of deterministic functions from Θ to R, where Θ is a compact subset of a Euclidean space (more generally Θ could be a compact subset of a metric space, and in particular is allowed to be a potentially infinite-dimensional space). Then $\{H_N\}$ converges uniformly to a function $H : \Theta \to R$ if and only iff a) $\{H_N\}$ converges pointwise to H, and b) $\{H_N\}$ is uniformly equicontinuous. Furthermore, H is necessarily a continuous function.*

Any standard textbook on functional analysis will contain a proof of Ascoli's Theorem. the proof is not difficult. We now consider a generalization of Ascoli's Theorem in the case where $\{H_N\}$ is a random sequence of functions. We now need to define what we mean by *strong stochastic equicontinuity:*

Definition: *Let $\{H_N\}$ be a random sequence of functions from Θ to R. We say that $\{H_N\}$ is weakly (uniformly) stochastically equicontinuous if*

$$\overline{\lim}_{N\to\infty} P\left\{ \sup_{\|\theta-\theta'\|\le\delta} |H_N(\theta) - H_N(\theta')| > \epsilon \right\} < \epsilon. \tag{15}$$

Definition: *Let $\{H_N\}$ be a random sequence of functions from Θ to R. We say that $\{H_N\}$ is strongly (uniformly) stochastically equicontinuous (SSE) if $\sup_N |H_N(\theta)| < +\infty$ with probability 1 for all $\theta \in \Theta$ and if the sequence of random functions $\{G_N\}$ is weakly stochastically equicontinuous, where G_N is defined by $G_N(\theta) \equiv \{\sup_{m\ge N} H_m(\theta)$.*

The following theorem can be viewed as the stochastic version of Ascoli's Theorem: it provides necessary and sufficient conditions for the strong uniform convergence of a sequence of random functions:

Theorem: *Let $\{H_N\}$ be a sequence of strongly uniformly stochastically equicontinuous functions from Θ to R, where Θ is a compact subset of a Euclidean space (more generally Θ could satisfy the weaker restriction of being* **totally bounded***). Then $\{H_N\}$ converges uniformly to a function $H : \Theta \to R$ with probability 1 if and only iff a) $\{H_N\}$ converges pointwise to H with probability 1, and b) $\{H_N\}$ is strongly uniformly stochastically equicontinuous.*

For a proof of this Theorem, see D. Andrews (1992) "Generic Uniform Convergence" *Economic Theory* **8**, 241–247.