University of Illinois — Department of Economics
Spring 2017 — Econ 574 — Roger Koenker

Lecture 3
"Everything that Rises Must Converge"

Often we are interested in sequences $X_1, X_2, \ldots$ of r.v.'s on a p-space $(\Omega, \mathcal{A}, P)$. It is fashionable now to speak of asymptopia, a mythical land which serves as a laboratory of statistics where we may conduct thought experiments to compare performance of estimation and inference procedures. In this land sample sizes tend to infinity and comparisons are much easier than the workaday world of everyday life. "Easier than what?" you may ask. Easier than exact finite sample results, I would say. Here we will emphasize validation by monte-carlo of these thought experiments and the constructive interplay between these two tools. We are interested in behavior of sequences of r.v.'s, and in particular, in the convergence of the tail elements of the sequence.

1. *Convergence in Probability*

   Let $\{X_i\}_{i=1,2\ldots}$ and X be real valued r.v.'s on $(\Omega, \mathcal{A}, P)$. We say X_n *converges in probability* to X if

   $$\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$$

   for any $\varepsilon > 0$. And we will write $X_n \to X$.

   Often X will be a degenerate r.v., e.g., If $X_n = \bar{X}_n = n^{-1}\sum_{i=1}^n Z_i$ where Z_i are iid $\mathcal{N}(\mu, \sigma^2)$ then $X_n \to \mu$, and we can think of μ as the degenerate r.v. X which takes the value $\mu \quad wp1$.

2. *Convergence with Probability 1*

   We say X_n converges $wp1$, or almost surely, to X if

   $$P(\lim_{n\to\infty} X_n = X) = 1$$

   or equivalently, for any $\varepsilon > 0$,

   $$\lim_{n\to\infty} P(|X_m - X| < \varepsilon, \text{ for all } m > n) = 1.$$

   Sensibly, almost sure convergence $\Rightarrow$ convergence in probability, but we will encounter examples in which the converse fails.

3. *Convergence in q^{th} Mean*

   X_n converges in q^{th} mean to X if

   $$\lim_{n\to\infty} E|X_n - X|^q = 0$$

By the moment inequality $X_n \stackrel{q^{\text{th}}}{\to} X \Rightarrow X_n \stackrel{p^{\text{th}}}{\to} X$ for any $0 < p < q$. Often $q = 2$, in practice. As an example of extreme behavior suppose that $X_n = 0$ with probability $1 - n^{-3}$ and $X_n = n$ with probability n^{-3}, then taking $X = 0$, we have $\lim E|X_n - X|^q = 0$ for $q = 1, 2$, but $E|X_n - X|^3 = 1$.

4. *Convergence in distribution (law)*

   X_n converges in distribution to X if for their respective distribution functions

$$\lim_{n\to\infty} F_n(x) = F(x) \qquad \text{at each continuity point of } F$$

   We will write this as $X_n \stackrel{\mathcal{D}}{\to} X$, or as $F_n \Rightarrow F$, the latter is generally pronounced F_n converges weakly to F. Often I'll just write $X_n \rightsquigarrow X$, or e.g. $X_n \rightsquigarrow \mathcal{N}(0,1)$.

5. *The Story of O*

   For positive deterministic sequences $\{a_n\}, \{b_n\}$

   1 If there is a $\Delta < \infty$ such that $a_n/b_n \le \Delta$ for sufficiently large n, we say

$$a_n = O(b_n)$$

   2 If $a_n/b_n \to 0$ we say

$$a_n = o(b_n)$$

   Clearly, if $a_n = O(n^r)$ and $b_n = O(n^s)$, then $a_n b_n = O(n^{r+s})$ and $a_n + b_n = O(n^{\max(r,s)})$ and similarly for o. This useful device was extended to r.v.'s by Mann and Wald (1943) and developed somewhat by Pratt and others.

   For sequences $\{X_n\}$ and $\{Y_n\}$ of r.v.'s on $(\Omega, \mathcal{A}, P)$ and any $\varepsilon > 0$,

   1' If there exists $\Delta < \infty$ such that $P(|X_n| \ge \Delta |Y_n|) < \varepsilon$ for n sufficiently large, we write $X_n = O_p(Y_n)$.

   2' If $P(|X_n| \ge \varepsilon |Y_n|) \to 0$, we write $X_n = o_p(Y_n)$

   Often, Y_n will be nonstochastic, and in particular we will often write

$$\begin{array}{ll} X_n = O_p(1) & \text{for “bounded in probability”} \\ X_n = o_p(1) & \text{for “tending to zero in probability”} \end{array}$$

   Further details on O_p and o_p are provided in the handout from Bishop, Fienberg and Holland(1975) *Discrete Multivariate Analysis*, which is available from the web site in the "Readings" section.

6. *Some Basic Tools*

   Thm: $X_n \stackrel{q^{\text{th}}}{\to} X \Rightarrow X_n \to X$.

   Pf: $E|X_n - X|^q \ge E[|X_n - X|^q I(|X_n - X| > \varepsilon) \ge \varepsilon^q P(|X_n - X| > \epsilon)$.

Thm (Prop 4.1 of Shorack (2000)). Suppose that $X \sim F, X_n \sim F_n$ such that $X_n \to X$. Then $X_n \to_d X$.

Pf:

$$\begin{aligned} F_n(t) = P(X_n \le t) &\le P(X \le t+\varepsilon) + P(|X_n - X| \ge \varepsilon) \\ &\le F(t+\varepsilon) + \varepsilon \qquad \text{for } n \ge n_\varepsilon \end{aligned}$$

And

$$\begin{aligned} F_n(t) = P(X_n \le t) &\ge P(X \le t-\varepsilon, |X_n - X| \le \varepsilon) \\ &\equiv P(A \cup B) \\ &\ge P(A) - P(B^C) \\ &= F(t-\varepsilon) - P(|X_n - X| \le \varepsilon) \\ &\ge F(t-\varepsilon) - \varepsilon \qquad \text{for } n \ge n'_\varepsilon \end{aligned}$$

Thus, for $n \ge \max\{n_\varepsilon, n'_\varepsilon\}$ we have

$$F(t-\varepsilon) - \varepsilon \le \underline{\lim} F_n(t) \le \overline{\lim} F_n(t) \le F(t+\varepsilon) + \varepsilon$$

and for all continuity points of F the result follows by letting $\varepsilon \to 0$. □

Thm: If $\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty$ for every $\varepsilon > 0, X_n \to X a.s.$

Pf:

$$\begin{aligned} P(|X_n - X) > \varepsilon \quad \text{for} \quad \text{some} \quad m > n) &= P(\cup_{m=n}^{\infty}\{|X_m - X| > \varepsilon\}) \\ &\le \sum_{m=n}^{\infty} P(|X_m - X) > \varepsilon) \end{aligned}$$

which converges to 0, by hypothesis.

Remark: This result illustrates the gap between $\to$ and $\to a.s.$, if $X_n \to X$ "sufficiently fast", then $X_n \to X a.s.$

Thm: For $\{X_n\}$ and X with df's $\{F_n\}, F$ and cf's $\{\phi_n\}, \phi$, the following are equivalent

(i) $F_n \Rightarrow F$

(ii) $\lim \phi_n(t) = \phi(t) \qquad t \in \mathbb{R}$.

(iii) $\lim \int g \; dF_n = \int g \; dF$ for each bounded continuous function g.

Pf. (i) $\Rightarrow$ (iii) is Helly Thm, converse follows by taking $g(x) = I(x < t) + (x-t)I(t < x < t+\varepsilon)$ where,

$$F(t-\varepsilon) \le \underline{\lim} F_n(t) \le \overline{\lim} F_n(t) \le F(t+\varepsilon)$$

for (ii) $\Leftrightarrow$ (iii), see Gnedenko (1962) §38.

Remark: The crucial implication of this is that if ϕ_n of X_n tends to $e^{-\frac{1}{2}t^2}$, then $X_n \rightsquigarrow \mathcal{N}(0,1)$.

Thm: (Cramér-Wold device) For $\{X_n\}$ and X in $\mathbb{R}^p$, $X_n \rightsquigarrow X$ iff for all $\alpha \in \mathbb{R}^p$, $\alpha' X_n \rightsquigarrow \alpha' X$.

Pf: Uses multivariate version of previous result.

Thm: (Slutsky) Let $X_n \rightsquigarrow X$ and $Y_n \to y$, a real constant. Then,

(i) $X_n + Y_n \rightsquigarrow X + y$

(ii) $X_n Y_n \rightsquigarrow yX$

Pf: We will prove (ii), (i) is similar. The argument given is that of Davidson; but note that the roles of X_n and Y_n are reversed there. For details on (i) see Bickel and Doksum, or Serfling. Suppose $y = 0$, for the convenience of the moment, and let $B > 0$, be a real constant, and denote

$$X_n^B = X_n I(|X_n| \leq B)$$

Then

$$\{|Y_n X_n| \geq \varepsilon\} = \{|Y_n||X_n^B| \geq \varepsilon\} \cup \{|Y_n||X_n - X_n^B| \geq \varepsilon\} \tag{1}$$

for any $\varepsilon > 0$,

$$\{|Y_n||X_n^B| \geq \varepsilon\} \subseteq \{|Y_n| \geq \varepsilon/B\}$$

and

$$P\{|Y_n||X_n^B| \geq \varepsilon\} \leq P\{|Y_n| \geq \varepsilon/B\} \to 0.$$

By hypothesis $X_n = O_p(1)$ so there exists $\delta > 0$, and $B_\delta < \infty$ such that for n sufficiently large,

$$P(|X_n - X_n^{B_\delta}| > 0) < \delta$$

Since

$$\{|Y_n||X_n - X_n^B| \geq \varepsilon\} \subseteq \{|X_n - X_n^B| > 0\}$$

so 1 and additivity implies,

$$\lim_{n\to\infty} P\{|X_n||Y_n| \geq \varepsilon\} < \delta$$

Since ε and δ were arbitrary we hae shown that $X_n Y_n \to 0$. The result follows by noting that Y_n can be replaced by $Y_n - y$.

Thm: (Continuous Mapping) If $X_n \rightsquigarrow X$ and g is continuous, $g(X_n) \rightsquigarrow g(X)$.

Pf: Follows immediately from weak convergence, but *extremely* useful.

Examples of the use of the CMT

(i) If $X_n \rightsquigarrow \mathcal{N}(0,1)$, then $X_n^2 \rightsquigarrow \chi_1^2$

(ii) If $(X_n, Y_n) \rightsquigarrow \mathcal{N}(0, I_2)$, then $X_n/Y_n \rightsquigarrow Z$, a standard Cauchy r.v.

(iii) When g isn't continuous beware!

$$g(t) = \begin{cases} t-1 & t \leq 0 \\ t+1 & t > 0 \end{cases}$$

Let $X_n = \frac{1}{n}$ $wp1$ and $X = 0$ $wp1$, so, $X_n \to X$, but $g(X_n) \to 1$ but $g(X) = g(0) = -1$. The conditions can actually be weakened slightly so that g can be discontinuous on a set of P-measure zero, but this isn't typically very helpful. See Resnick, p260 for details on this more general version.

Thm: (δ-method) Suppose $a_n(X_n - b) \rightsquigarrow X$ where a_n is a sequence of constants tending to ∞, and b is a fixed number. Let $g : \mathbb{R} \to \mathbb{R}$ be differentiable with continuous derivative g' at b. Then

$$a_n(g(X_n) - g(b)) \rightsquigarrow g'(b)X.$$

Pf: By Slutsky,

$$X_n - b = a_n^{-1}[a_n(X_n - b)] \to 0$$

and therefore $X_n \to b$. Now apply mean value theorem to $g(X_n) - g(b)$,

$$g(X_n) - g(b) = g'(X_n^*)(X_n - b)$$

where $|X_n^* - b| \leq |X_n - b|$,whence $X_n^* \to b$ so by the continuity of g' and the CMT $g'(X_n^*) \to g'(b)$. Multiplying by a_n and again applying Slutsky we have the result. The same argument generalizes to $X_n, X \in \mathbb{R}^p$.